20 years Mathematical Optimization in the Real World
or
how to Survive Industry Projects as a Mathematician

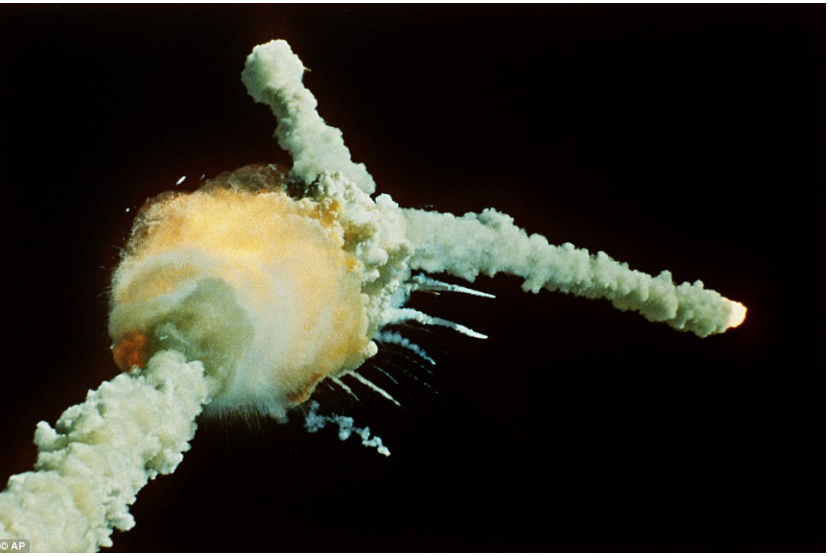
Thorsten Koch

Zuse Institute Berlin

Technische Universität Berlin

Combinatorial Optimization @ Work
Online September 2020

Research



- What does **Research** mean?
- It means the outcome is not yet determined.
- A research project can succeed by showing that it is not possible to achieve the initial goal.
- Research is conducted by systematically trying untested paths and devising new methods.

Getting the project

How to convince the industry people that you can help them:

- They are the specialists for the topic not you.
- Be aware they do not want a result that says:
  We can prove there exists a unique solution.
- Even if you know something about their business, regardless whom you ask, they will tell you : "We are special"
  Corollary: Since everybody is special, they are all equal.
- If you try to convince them by showing something similar, they might have a very narrow view with little abstraction ability.
- If asked, how much you can improve on the current solution, the correct answer is 15%
  (see G. Dueck, DMV-Mitteilungen, 2003, 44-45)

The improvement potential is always 15%

- 5% ⇨ “So much we save by simply pushing the employees.”
- 10% ⇨ “Sounds poor. We could do similar ourselves if we would get as much money as you ask for.”
- 20% ⇨ “this sounds very ambitious. You must remember: if we give you the money, we have to promise 20% to our boss. We dare not to do this.“
- 30% ⇨ “Braggart! Get out!”

From this it follows that I have to say 15%.

- I said 15% and immediately got a signature
- I said 13 % ⇨ “Why such a crooked number? How could you be so precise?”
- I said 14%, same result.

I stayed at 15 percent. Always 15 percent. Only 15 percent. All nodded, everybody satisfied. I had discovered an absolute Natural constant!

Mathematics always saves 15%. Completely regardless of the Problem!

15 Prozent. QED. Oder gibt es schon falsche Fünfzehner?! Gunter Dueck. Das Sintflutprinzip. Springer. 2000

Pilot projects

Sometimes a company will suggest to do a pilot project first:

- The unspoken expectation is that you put in more resources than what you are paid for.
- Chances for a continuation project are as good with or without a pilot project.
- You will have trouble to get up your prices again afterwards.
- If you do this, the default has to be the continuation. Just suggest the right to drop out at a certain point in case of failure.

Making a contract

- Deliverables
- Intellectual property rights (patents)
- 3rd party code
- Don't do maintenance
- Right to publish
- Right to give talks
- Right to cooperate with others (incl. NDA conditions)
- Right to continue afterwards with others (competitors)
- Rights on data (esp. afterwards)
- What to do if the industry partner does not keep their milestones?

Remember: The contract is basically useless, as you will never sue and can do little later on.

Legal stuff

The sublicensing right described in section 4.2 above shall include the right to a) sublicense the Software to a Financier for sublicensing to End Users or b) permit End Users to sublicense the Software to a Financier. The rights described in a) and b) shall only be granted in order to enable a financing model for the Licensee and the Software for End Users. To allow for financing of the Software, a Financier may lease, rent, sublicense or temporary transfer the Software for temporary use to an End User during the financing term. At the end of the financing term, a Financier may transfer or offer to sell the license for the Software to the End User with whom Financier had agreed upon financing of the Software. The standard license terms must state explicitly that any on-sale of the Software by a Financier to a third party who is different from the End User who financed the Software is subject to the prior written approval of Licensee and Licensor.

Administrative stuff

Is your product subject to the U.S. Export Administration Regulations (EAR), 15 CFR Parts 730-774?

Does your product fall under the U.S. Government International Traffic in Arms Regulations (ITAR) or any other local government regulations which uniquely control the product for military or encryption reasons?

Please indicate any available classification under EU Law (Council Regulation (EC) No 428/2009 of 27 August 2009 setting up a Community regime for the control of exports of dual use items and technology)

a. ECCN:

b. Does the General Software Note apply?

c. Comments:

CCATS : Commodity classification automated tracking system, alpha-numerical code assigned by US BIS.

Do you have the heart to sell software?

Licensor represents and warrants that it is the owner of the Software Products, including all intellectual property rights there under, copyright, patent, trademark, trade secret and other applicable law.

Licensor shall fully indemnify, hold harmless and defend Licensee against law suits based on any claim that the Software Products infringe any patent, copyright, trademark, trade secrets, or other proprietary right.

Licensor remains liable to fully compensate the entity concerned for any damages, costs and reasonable expenses incurred in connection with such third party intellectual property infringement claims.

One option is only to provide a method that afterwards is implemented (again) by a commercial software developer.

Lawyers

The company will involve lawyers.
Your institute or university might do also:

- Lawyer like to dispute by attrition and exhaustion.
- They often do not understand what the project is really about.
- They have no problem to argue at length about how to distribute the members of the empty set.
- They are obsessed with low probability worst cases.
- They try to cover all cases without any formal method and often without understanding the concept.
- They will usually not converge unless by massive time pressure and order from above.
- They are necessary and it will take time.
- You do not understand the implications of their writing.

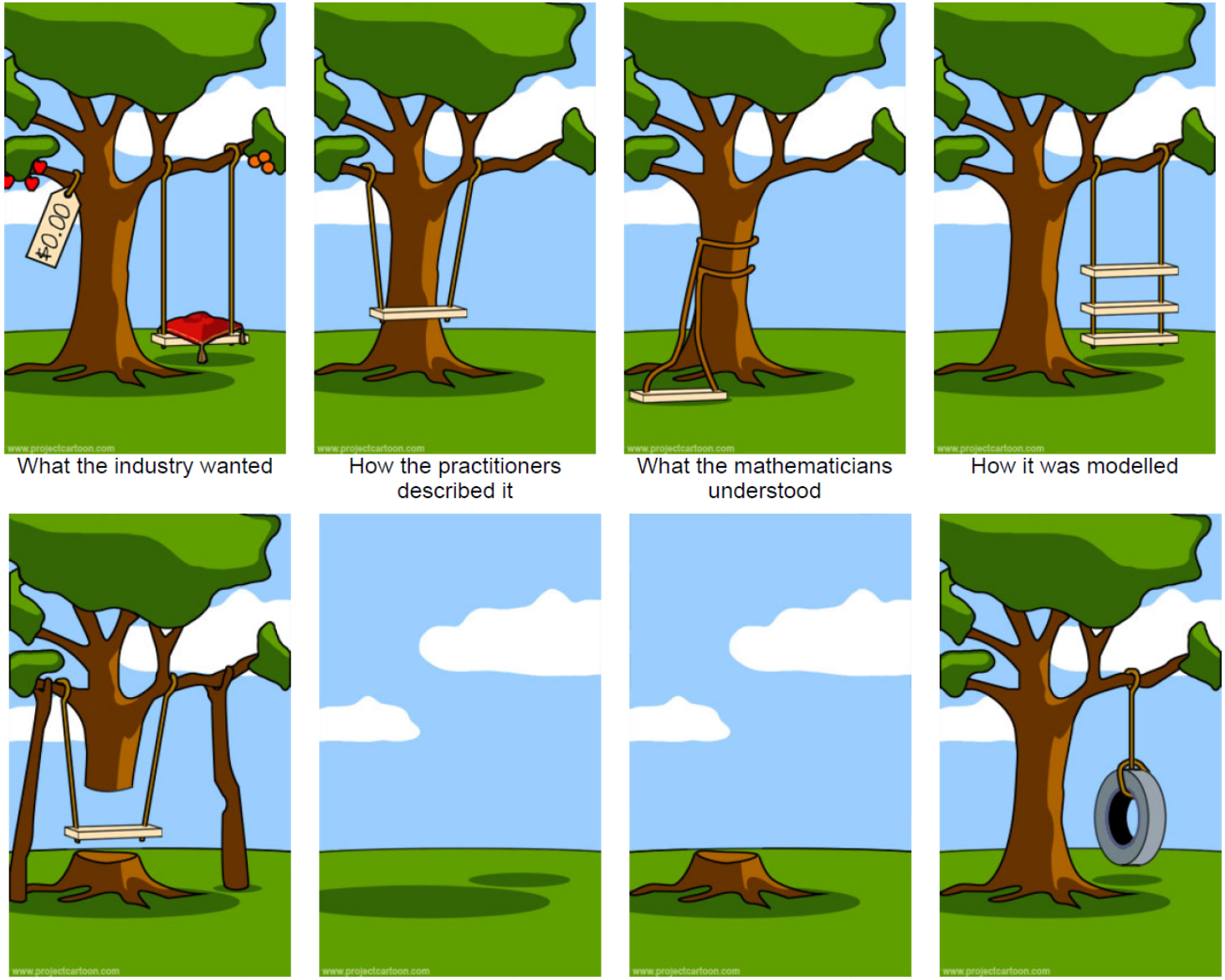
Communication

The same words may have different meanings in different Communities:

Speak the language of the problem owner

- Technical terms
- Mother tongue
- Their problem is your problem,
  and your solution has to become their solution.
- Do not trust assumptions.
- Convince the decision makers – not only the techies.

Describing the problem



What the industry wanted

How the practitioners described it

What the mathematicians understood

How it was modelled

How it was implemented

How the project was documented

How it was supported

What was really needed

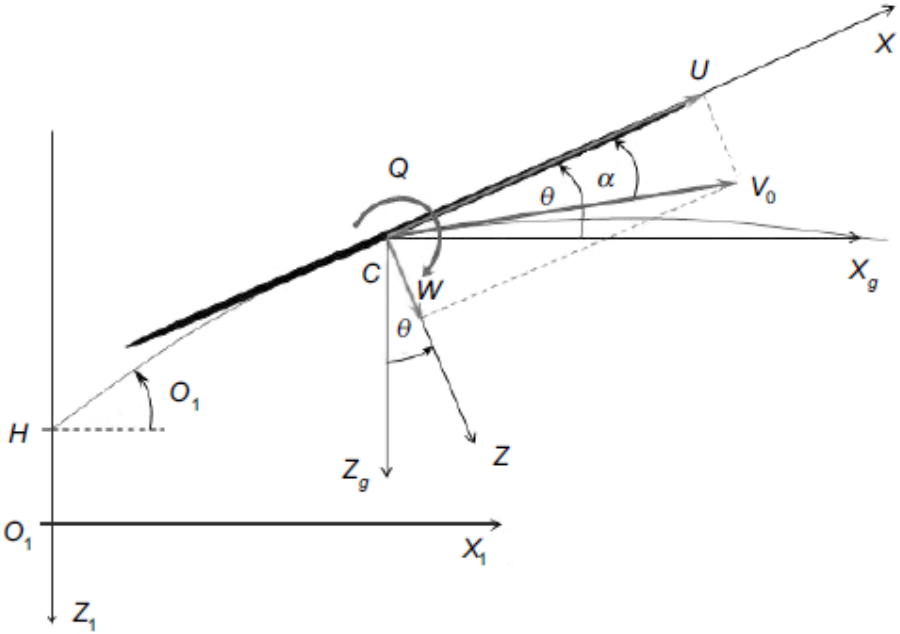
Modelling or what is the objective?

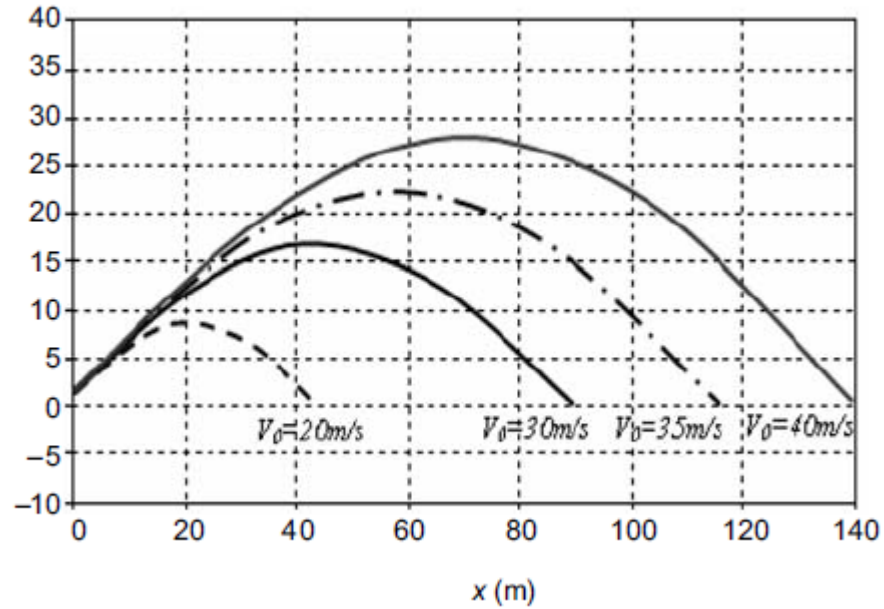
- Find out what the real objective is.
  Usually it is not what you are told in the beginning.
- Sometimes it is difficult because the cost impact when optimizing virtual or already existing structures is unclear.
- Often you have to compare apples and pies.
- Preprocessing is important. Most real-world problems are rather big, at least compared to “academic toy examples”.
- Decide what you have to model and what to ignore (or fix later)

The traveling salesman problem is to mathematical programming what chess is to artificial intelligence: thoroughly useless and fiercely competitive sport that serves as a testing ground of your techniques.

—Vasek Chvatal

Javelin throwing: Theory



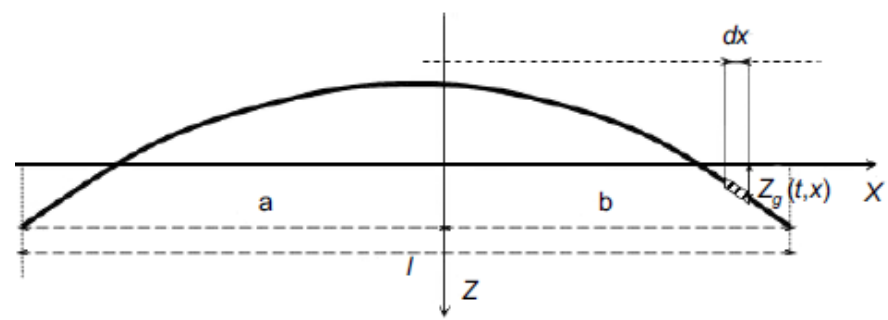


$$\frac{d}{dt}\left(\frac{\partial T^*}{\partial U}\right)+\frac{\partial T^*}{\partial W}Q=Q_U^*$$

$$\frac{d}{dt}\left(\frac{\partial T^*}{\partial W}\right)+\frac{\partial T^*}{\partial U}Q=Q_w^*$$

$$\frac{d}{dt}\left(\frac{\partial T^*}{\partial Q}\right)+\frac{\partial T^*}{\partial U}W-\frac{\partial T^*}{\partial W}U=Q_Q^*$$

$$\frac{d}{dt}\left(\frac{\partial T^*}{\partial q}\right)-\frac{\partial T^*}{\partial q}+\frac{\partial V_{zg}}{\partial q}U=Q_q^*$$



Mathematical Modeling and Numerical Simulations of Javelin Throw
J. Maryniak, E. Ładyżyńska-Kozdraś, E. Golihska
J. of Human Movement, Vol. 1 (2009), 16-20

Javelin throwing: Sports

Javelin throwing: Real-life

Modelling: How strange do we want to get

Do you know what
Zone defining Origins (verzonende Ursprünge) or
a Breathing Sack (atmender Sack) are?

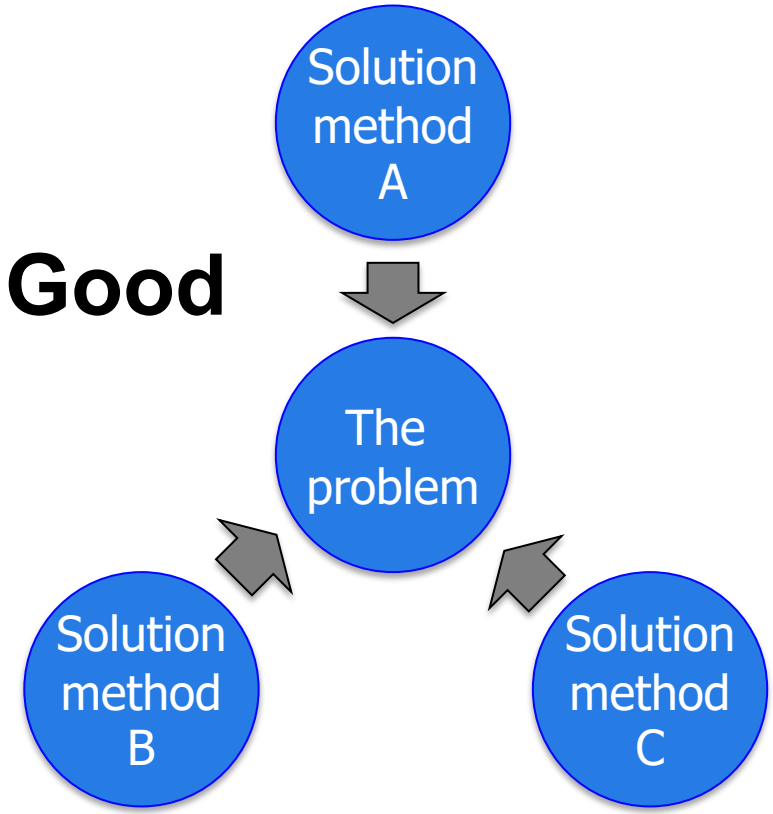
Industry partner: “We have one location where there is still this old XXX which behaves totally different.”

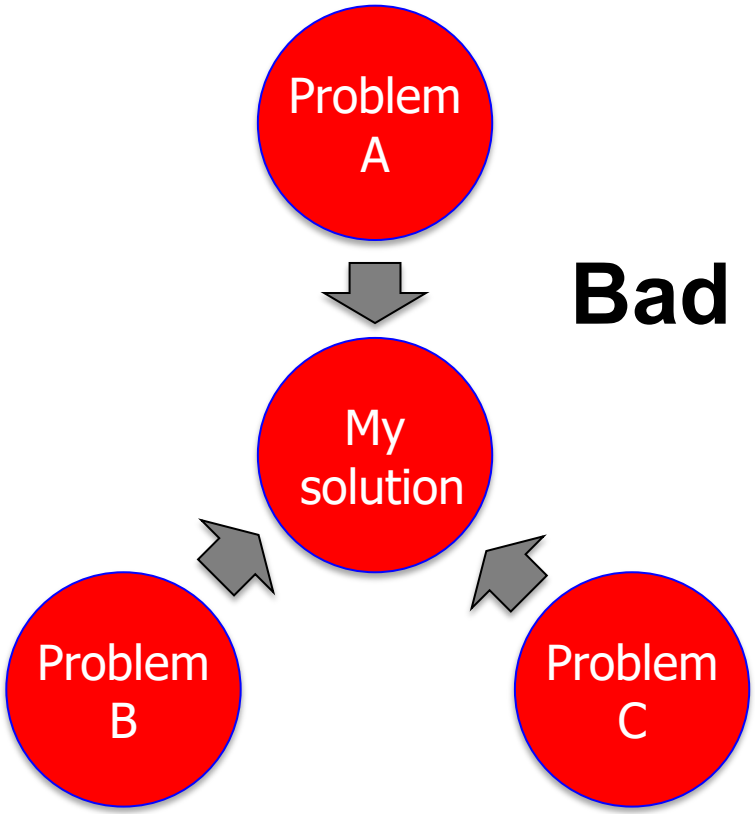
Problem:

Decide what you have to model and what to ignore
(or fix later)

Think and act problem centric

- It is more important to solve the right problem than to solve the problem right.
- Identifying the problem is half of the way to the solution.





The project setup / classical approach

We have a problem to solve,

The project setup / classical approach

we have a teacher,

The project setup / classical approach

… and we have a very determined PhD student.

Now, the student supervised by the teacher attacks the problem.

This is what we call the classical “Hero Approach”.

The project team

What if the problem is too big and you need a whole team to tackle it?

Maybe you do not have the necessary expertise and need to cooperate with other institutions.

Mathematical research usually has no suitable infrastructure to run big projects with non-disjunctive tasks.

Data

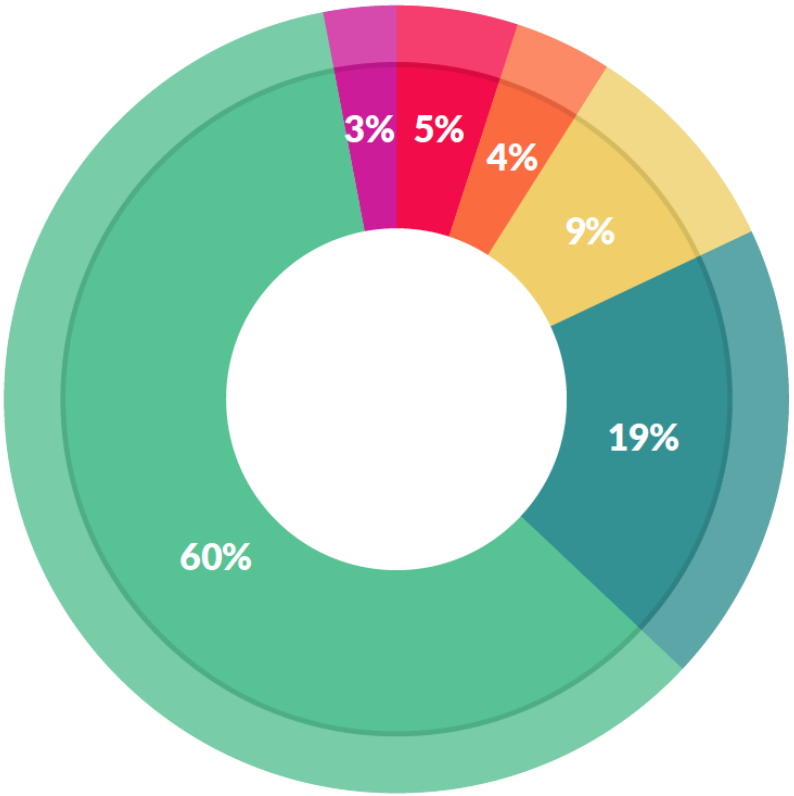
You would think a …

- … cellular network operator knows where its base stations are located?
- … fixed network operator can tell where the parts of its network are connected?
- … chemical company knows how many plants they have?
- … 5 m long pipeline cannot have a height difference from end-to-end of 100 m?

- Many companies have their data in Excel.
  There is no formal validation or referential integrality check.
- If they did formal validation, usually they found there was information they needed which they could not input, and they started to “reuse” some data fields.
- If there is not at least 1 error per 100 data sets you are not looking hard enough.
- Usually the data changes all the time.
- They might not want to give it to you.
- The data might just not exist.

The first result of an optimization project is usually to improve the quality of planning data available at the company.

Data Scientists spend time on ...



What data scientists spend the most time doing

- *Building training sets: 3%*
- *Cleaning and organizing data: 60%*
- *Collecting data sets; 19%*
- *Mining data for patterns: 9%*
- *Refining algorithms: 4%*
- *Other: 5%*

Cleaning Data

from Cloud Flowers Data Science Report 2016

`http://visit.crowdflower.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf`

Solutions

Optimality is often not required. Industry is usually more interested in reasonably good solutions in short time than in proven optimal solutions after a long wait.

To compute a gap or prove infeasibility is important.

What does it mean? You are provably global optimal but the former solution the company used is better.

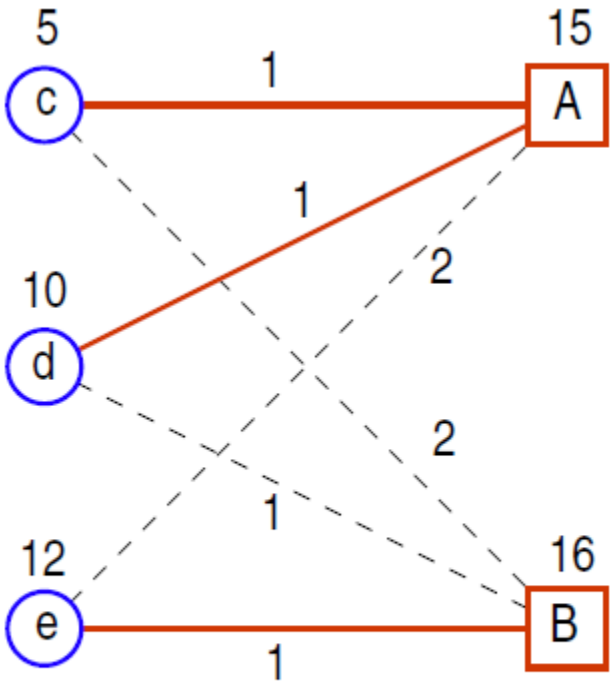
It is very important that the solution is 2-optimal.

Extremal Solutions vs. 80%

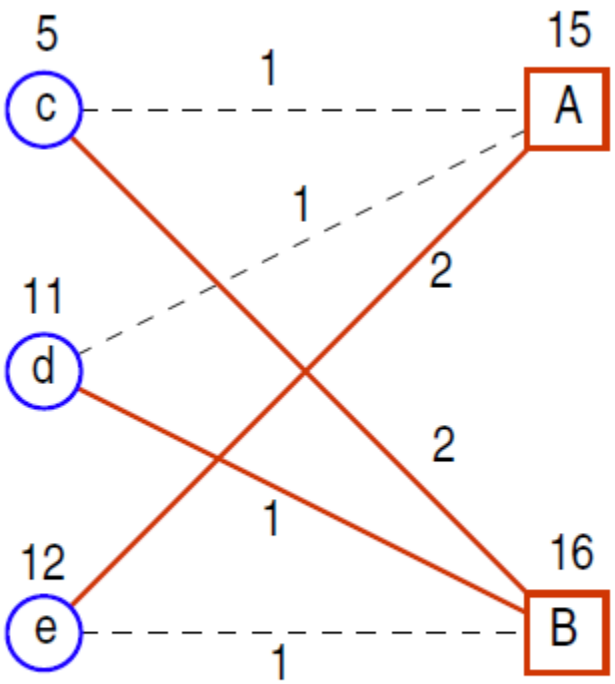
Strange solutions may have several reasons:

- Anomalies (errors) in the data
- Differences between model and reality
- Reaching of capacity or cost thresholds
- The result is just different than expected

Instable solutions



(a) Optimal solution



(b) After a small change

Strange cost functions

Cost per channel	Distance in km <45	<90	⩾90
uv to vv	3.30	4.95	6.60
vv to hv	3.50	5.25	7.00

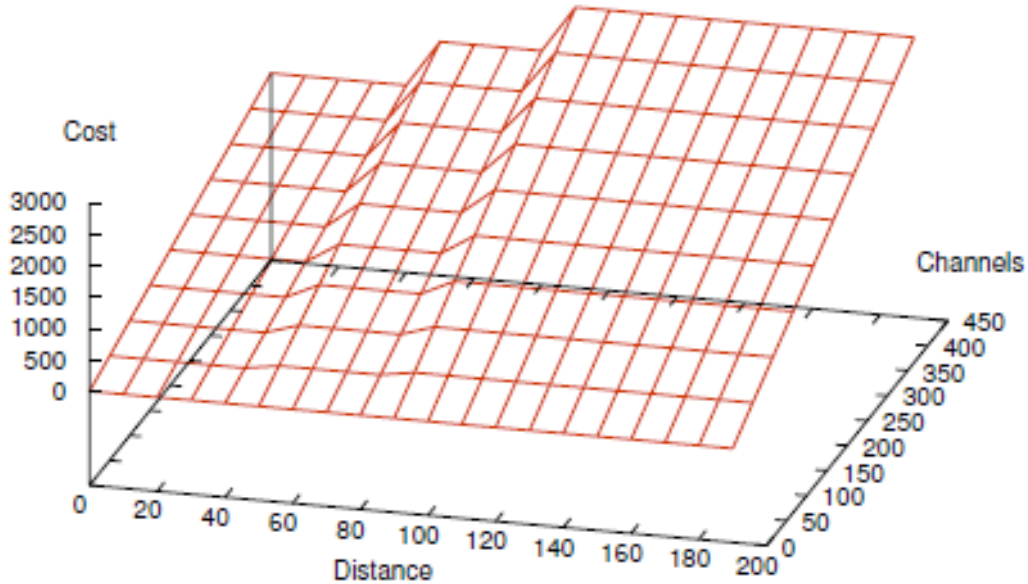
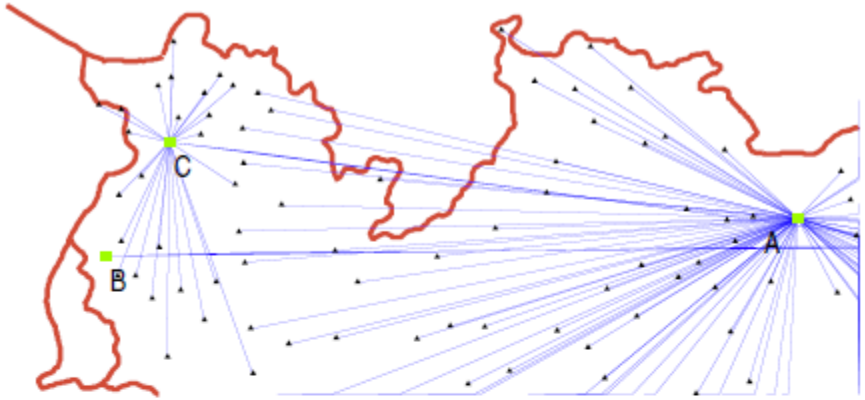


Figure 4.9: Cost function depending on distance and channels

Strange solutions



(a) Bee-line distances

(b) Transport network distances

- Why is no UV connected to B?

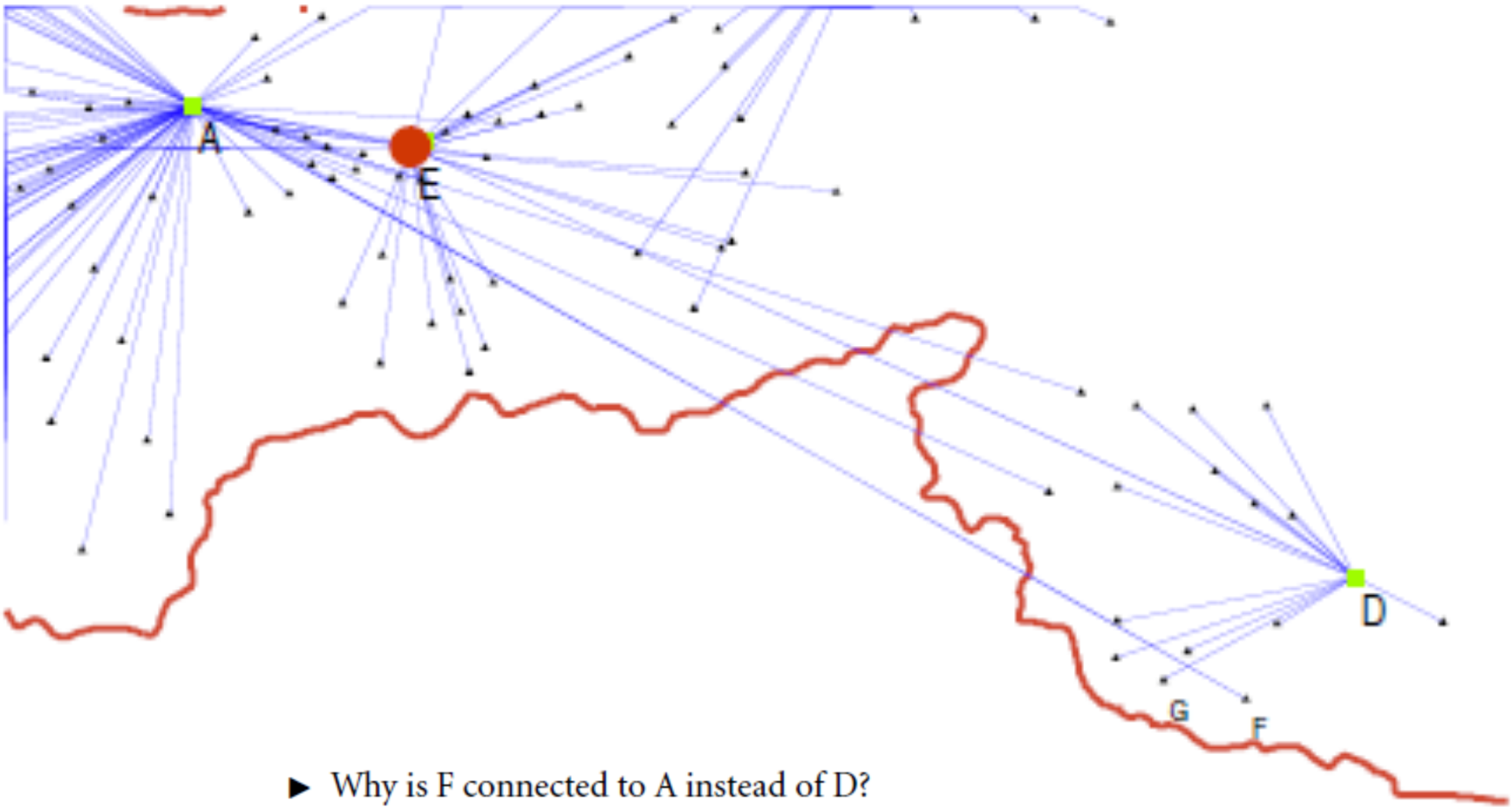
  Since all Uvs in question are less than 45km away from B and C, the connection costs are equal. Since C has enough capacity all the UV just happen to be connected to it.

- Why then are B and C both active?
  Because the input data requires it.

- Why are some UVs in the vicinity of C connected to A?

  Because connecting them to C would increase the total length of the link from the UV to the HV. VV to HV connections are only a little cheaper than UV to VV links. So the cost for first connecting to the more remote C does not pay of. This changes if instead of bee-line distances transport network distances are used.

Strange solutions



- Why is F connected to A instead of D?
- Would not E be better than A to connect F?
- Why is G connected to D, if F is not?

Run of the project

- The bigger the company, the more unstable is the department.
- If the project takes too long, they may lose interest.
- This is research not development.
- Milestones in the beginning are hard to meet for research projects.
- The original solutions of the company are often infeasible.
- Check the solutions, we do errors, too.

The way to real-world applications

	Problem definition	Real world constraints	Data	Code
Pure research	None	None	None	None
Applicable research	General	Unknown	Random/Simplified	Whatever
Applied research	General	Maybe	Random/Simplified	Whatever
Case study	Simplified	Some	Simplified	Whatever
Planning application	Simplified	Some more	Simplified/Real	Production
Control application	Complete	all	Real	24/7

Slide from Guido Sand

How to make optimization solutions work in industrial practice?
Have the right people with the right mindset!

- Industry is full of optimization problems, but they are often not obvious – identifying them is part of the job.
- Excellent mathematics which fits to the challenges of the application is necessary but not sufficient for success.
- Having the right people with the right mindset is a key to success.

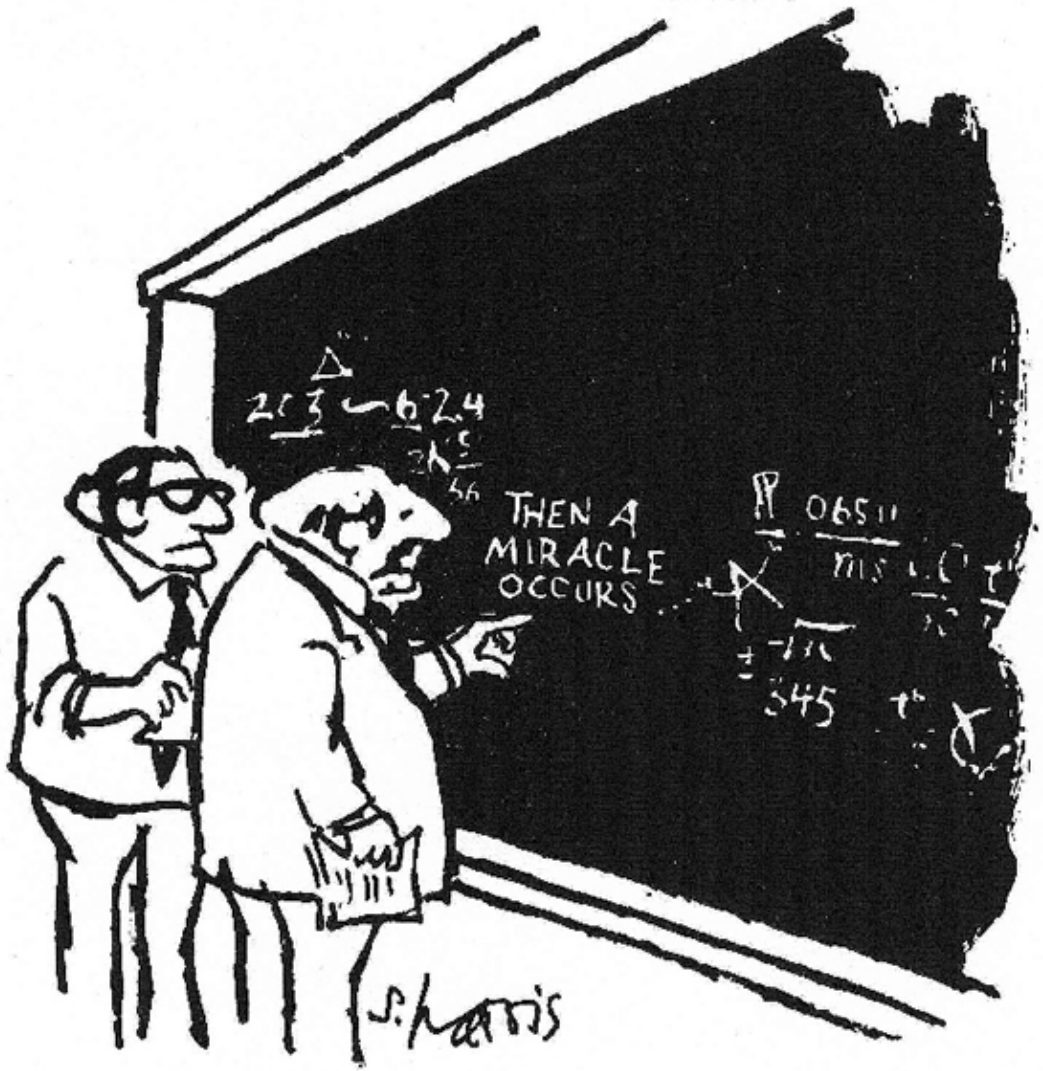
Why isn't it considered innovative,
if a solution works in industrial practice?

The final test of a theory is its capacity to solve the problems which originated it.

George Dantzig (1963) in

Linear Programming and Extensions

Thank you very much!



"I think you should be more explicit here in step two."