Part 1

Examples of optimization problems

Wolfgang Bangerth

What is an optimization problem?

Mathematically speaking:

Let X be a Banach space (e.g., R^n); let

$$f: X \rightarrow R \cup \{+\infty\}$$
$$g: X \rightarrow R^{ne}$$
$$h: X \rightarrow R^{ni}$$

be functions on X, find $x \in X$ so that

$$\begin{aligned} f(x) &\rightarrow \min! \\ g(x) &= 0 \\ h(x) &\geq 0 \end{aligned}$$

Questions: Under what conditions on *X, f, g, h* can we guarantee that (i) there is a solution; (ii) the solution is unique; (iii) the solution is stable.

What is an optimization problem?

In practice:

- *x={u,y}* is a set of design and auxiliary variables that completely describe a physical, chemical, economical model;
- *f(x)* is an objective function with which we measure how *good* a design is;
- *g(x)* describes relationships that have to be met exactly (for example the relationship between *y* and *u*)
- *h(x)* describes conditions that must not be exceeded

Then find me that *x* for which

$$
\begin{aligned}
f(x) &\rightarrow \min! \\
g(x) &= 0 \\
h(x) &\geq 0
\end{aligned}
$$

Question: How do I find this *x*?

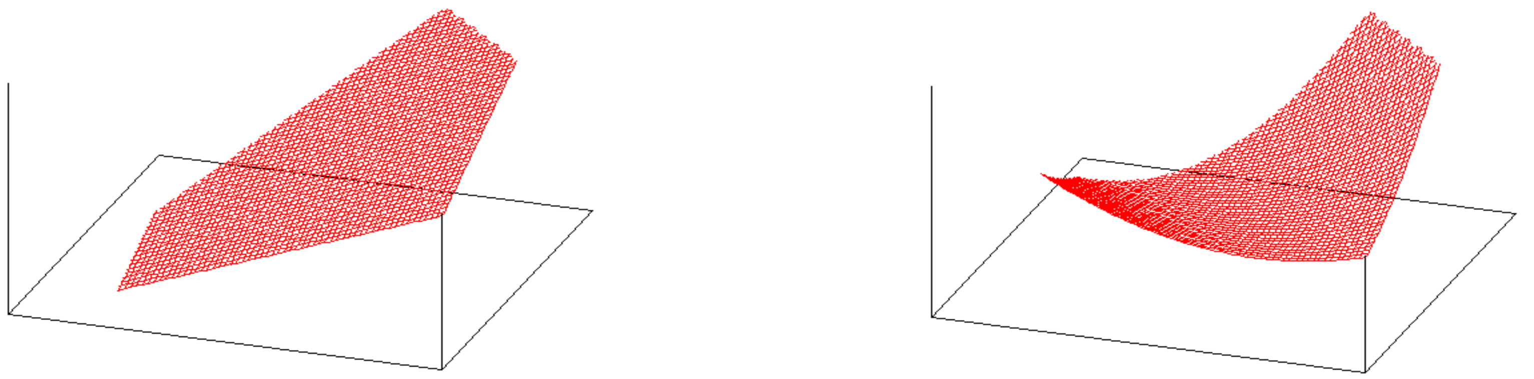
What is an optimization problem?

Optimization problems are often subdivided into classes:

Linear	vs.	Nonlinear
Convex	vs.	Nonconvex
Unconstrained	vs.	Constrained
Smooth	vs.	Nonsmooth
With derivatives	vs.	Derivativefree
Continuous	vs.	Discrete
Algebraic	vs.	ODE/PDE

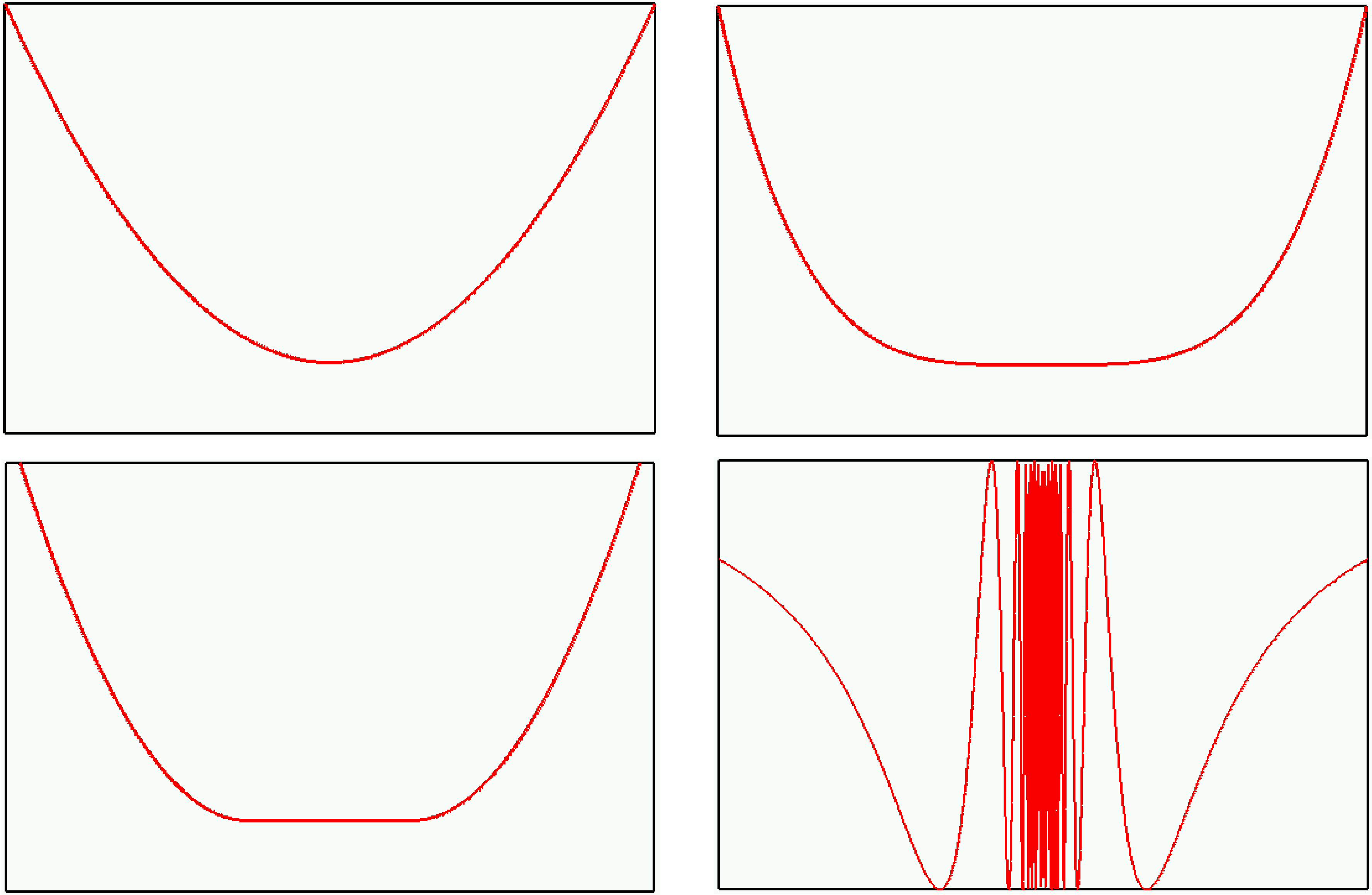
Depending on which class an actual problem falls into, there are different classes of algorithms.

Examples



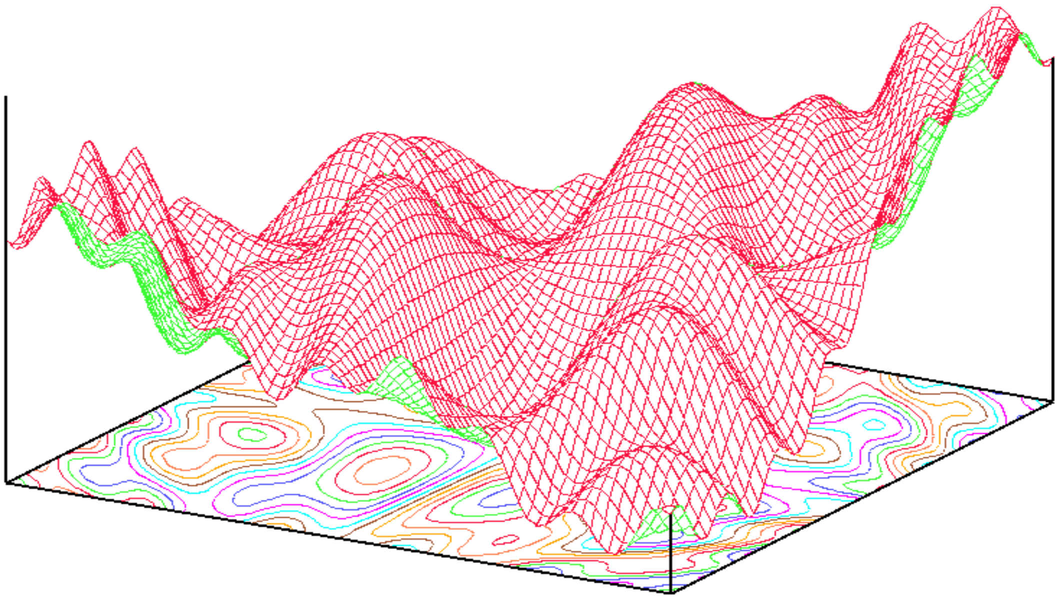
Linear and nonlinear functions *f(x)*
on a domain bounded by linear inequalities

Examples



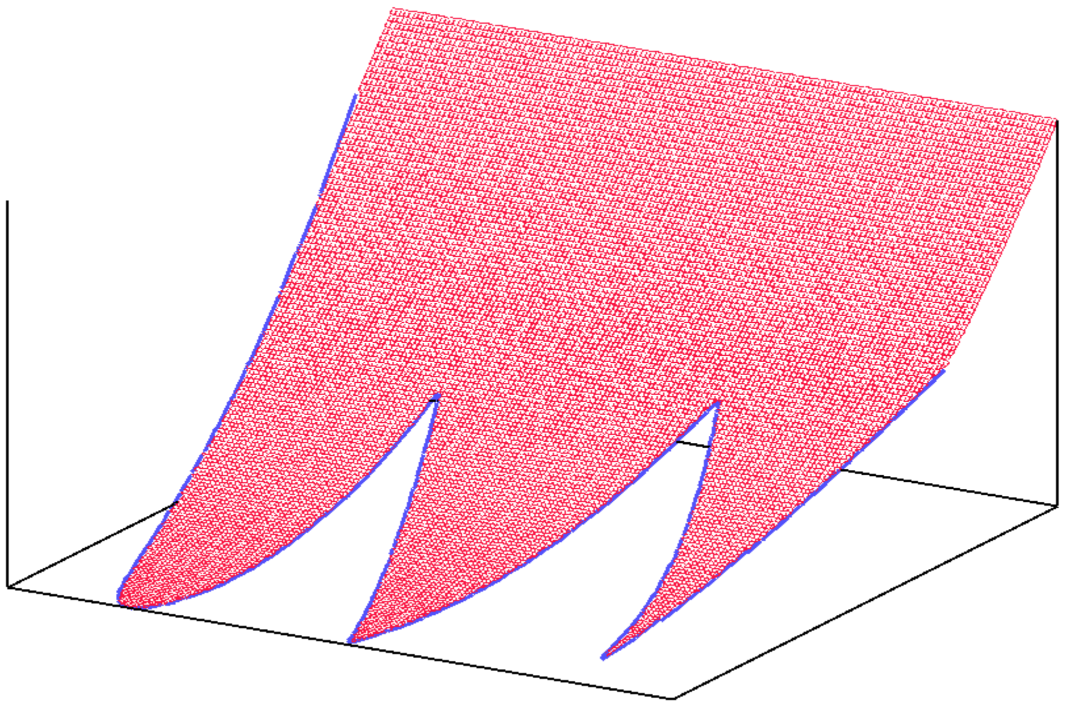
Strictly convex, convex, and nonconvex functions *f(x)*

Examples



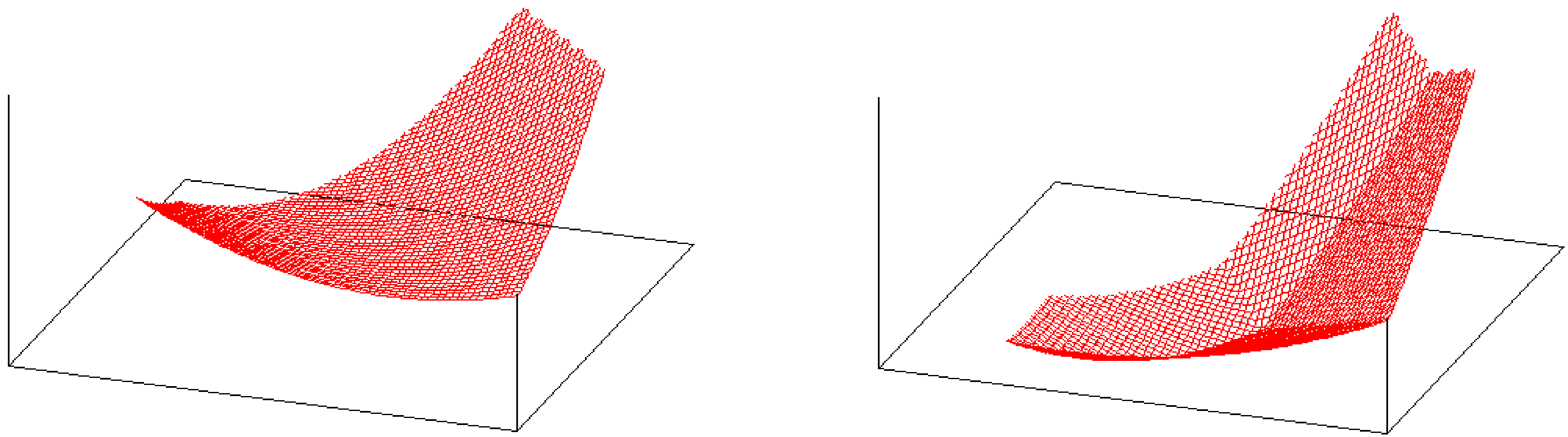
Another non-convex function with many (local) optima.
We may want to find the one *global* optimum.

Examples



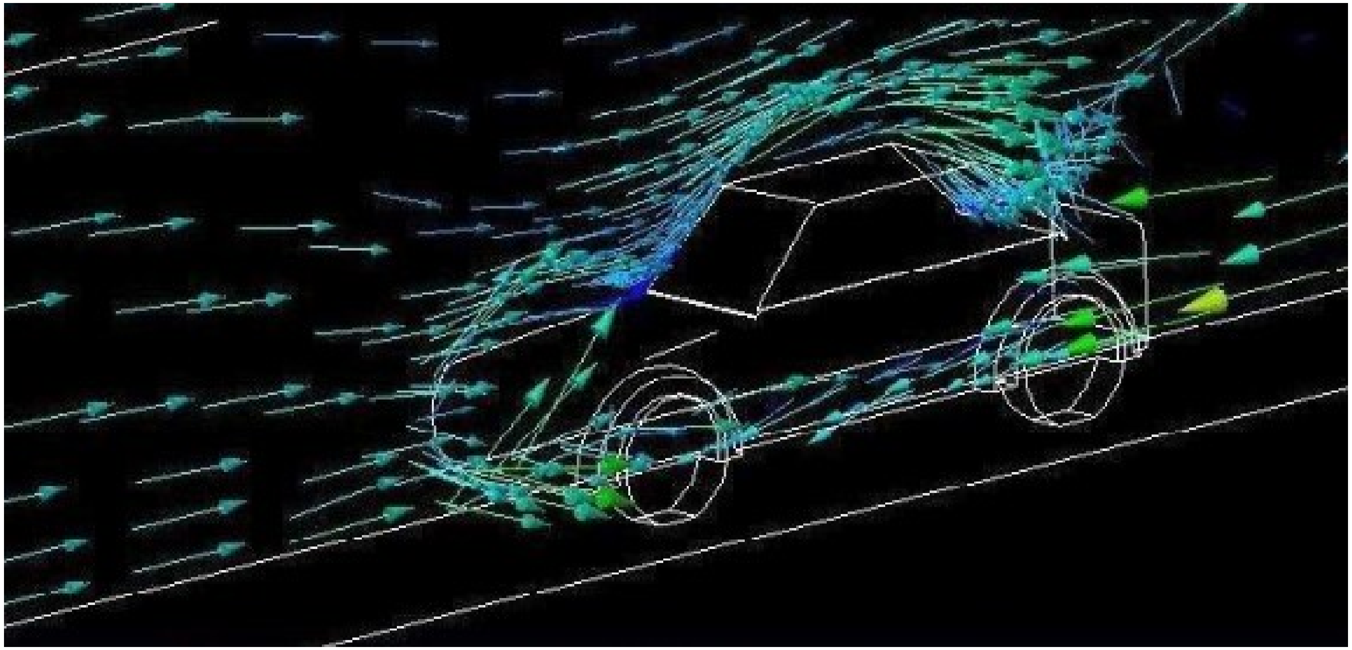
Optima in the presence of (nonsmooth) constraints.

Examples



Smooth and non-smooth nonlinear functions.

Applications: The drag coefficient of a car



Mathematical description:

$x=\{u,y\}$ — u are the design parameters (e.g. the *shape* of the car)
y is the flow field around the car

$f(x)$: the drag force that results from the flow field

$g(x)=y-q(u)=0$

constraints that come from the fact that there is a flow field $y=q(u)$ for each design. y may, for example, satisfy the Navier-Stokes equations

Applications: The drag coefficient of a car

Inequality constraints:

(expected sales price – profit margin) - cost(u) ≥ 0

volume(u) – volume(me, my wife, and her bags) ≥ 0

material stiffness * safety factor
- max(forces exerted by y on the frame) ≥ 0

legal margins(u) ≥ 0

Applications: The drag coefficient of a car

Analysis:

linearity:	*f(x)* may be linear *g(x)* is certainly nonlinear (Navier-Stokes equations) *h(x)* may be nonlinear
convexity:	??
constrained:	yes
smooth:	*f(x)* yes *g(x)* yes *h(x)* some yes, some no
derivatives:	available, but probably hard to compute in practice
continuous:	yes, not discrete
ODE/PDE:	yes, not just algebraic

Applications: The drag coefficient of a car

Remark:

In the formulation as shown, the objective function was of the form

$$f(x) = c_d(y)$$

In practice, one often is willing to trade efficiency for cost, i.e. we are willing to accept a slightly higher drag coefficient if the cost is smaller. This leads to objective functions of the form

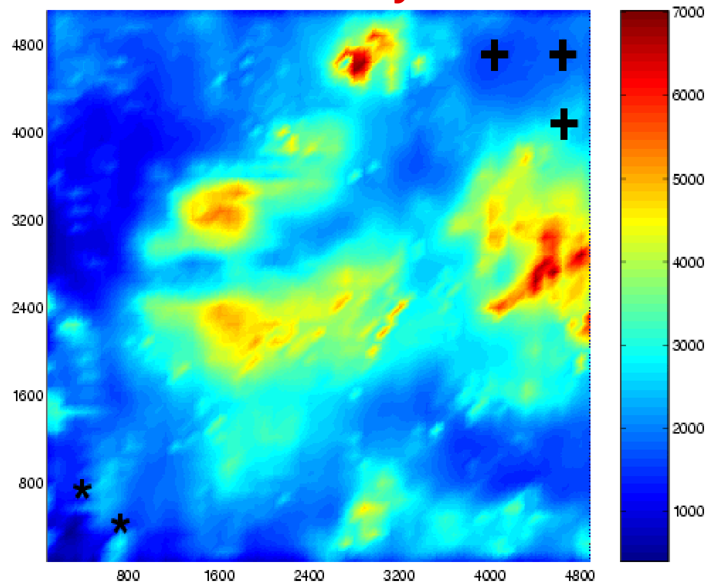
$$f(x) = c_d(y) + a\,\mathrm{cost}(u)$$

or

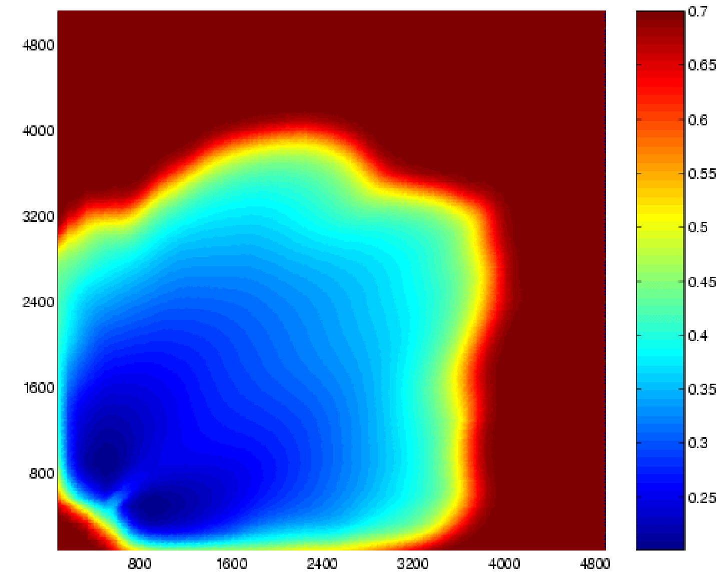
$$f(x) = c_d(y) + a[\mathrm{cost}(u)]^2$$

Applications: Optimal oil production strategies

Permeability field



Oil saturation



Mathematical description:

x={u,y} — *u* are the pumping rates at injection/production wells; *y* is the flow field (pressures/velocities)

f(x) — the cost of production and injection minus sales price of oil integrated over lifetime of the reservoir

g(x)=y-q(u)=0 — constraints that come from the fact that there is a flow field *y=q(u)* for each *u*. *y* may, for example, satisfy the multiphase porous media flow equations

Applications: Optimal oil production strategies

Inequality constraints *h(x)≥0*:

$U_{imax} - u_i \geq 0$ (for all wells *i*):

Pumps have a maximal pumping rate/pressure

produced_oil(*T*)/available_oil(*0*) – $c \geq 0$:

Legislative requirement to produce at least
a certain fraction

c_w - water_cut(*t*) ≥ 0 (for all times *t*):

It is inefficient to produce too much water

pressure – $d \geq 0$ (for all times and locations):

Keeps the reservoir from collapsing

Applications: Optimal oil production strategies

Analysis:

linearity:	*f(x)* is nonlinear *g(x)* is certainly nonlinear *h(x)* may be nonlinear
convexity:	no
constrained:	yes
smooth:	*f(x)* yes *g(x)* yes *h(x)* yes
derivatives:	available, but probably hard to compute in practice
continuous:	yes, not discrete
ODE/PDE:	yes, not just algebraic

Applications: Switching lights at an intersection

Mathematical description:

$x=\{T, t_i^1, t_i^2\}$ round-trip time *T* for the stop light system,

switch-green and switch-red times for all lights *i*

$f(x)$ number of cars that can pass the intersection per hour; to be maximized.

Note: unknown as a function, but we can measure it

Applications: Switching lights at an intersection

Inequality constraints $h(x)\geq 0$:

$300 - T \geq 0$:

No more than 5 minutes of round-trip time, so that people don't have to wait for too long

$t_i^2 - t_i^1 - 5 \geq 0$:

At least 5 seconds of green at each light *i*

$t^1_{i+1} - t_i^2 - 5 \geq 0$:

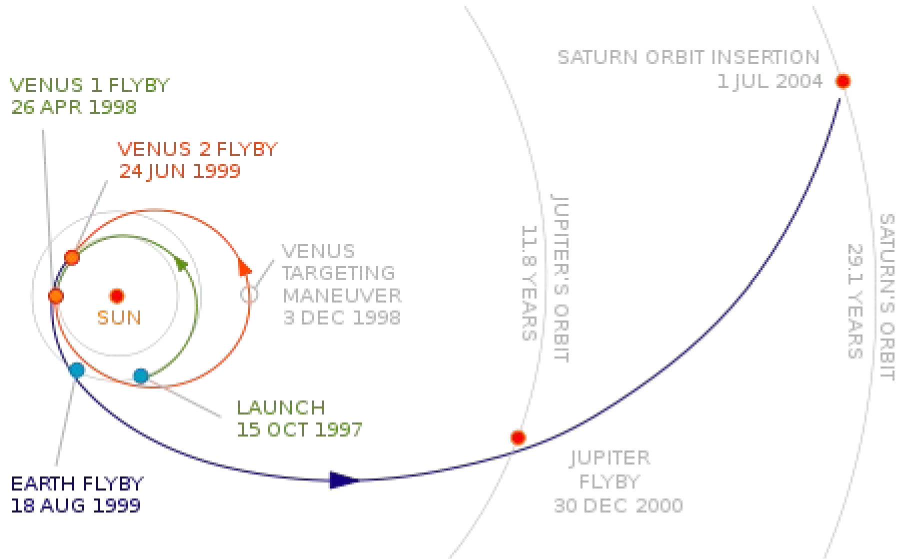
At least 5 seconds of all-red between different greens

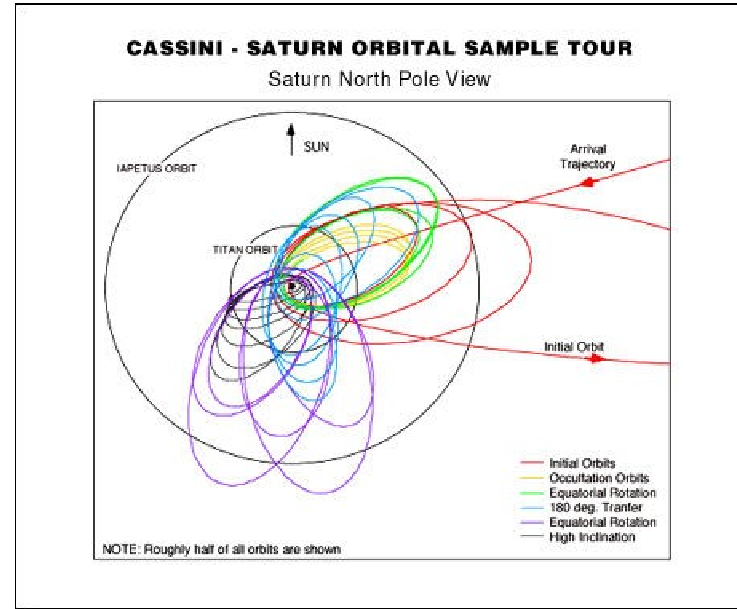
Applications: Switching lights at an intersection

Analysis:

linearity:	*f(x)* ?? *h(x)* is linear
convexity:	??
constrained:	yes
smooth:	*f(x) ??* *h(x)* yes
derivatives:	not available
continuous:	yes, not discrete
ODE/PDE:	no

Applications: Trajectory planning





Mathematical description:

$x = \{y(t), u(t)\}$ position of spacecraft and thrust vector at time t

$f(x) = \int_0^T |u(t)|\, dt$ minimize fuel consumption

$m\,\ddot{y}(t) - u(t) = 0$ Newton's law

$|y(t)| - d_0 \geq 0$ Do not get too close to the sun

$u_{\max} - |u(t)| \geq 0$ Only limited thrust available

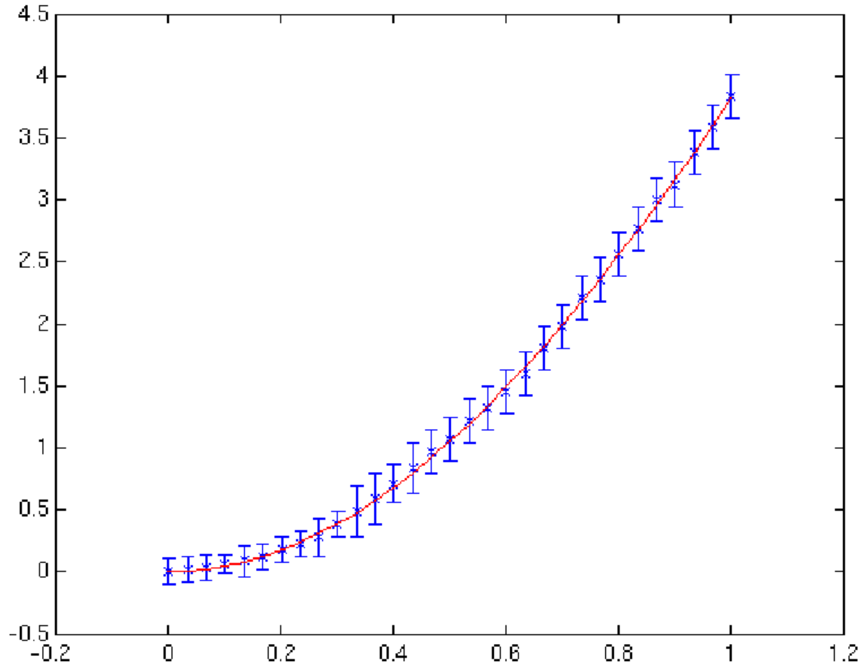
Applications: Trajectory planning

Analysis:

linearity:	*f(x)* is nonlinear *g(x)* is linear *h(x)* is nonlinear
convexity:	no
constrained:	yes
smooth:	yes, here
derivatives:	computable
continuous:	yes, not discrete
ODE/PDE:	yes

Note: Trajectory planning problems are often called *optimal control*.

Applications: Data fitting 1



Mathematical description:

x={a,b} parameters for the model $y(t)=\frac{1}{a}\log\cosh(\sqrt{ab}\,t)$

$f(x)=1/N \sum_i |y_i-y(t_i)|^2$

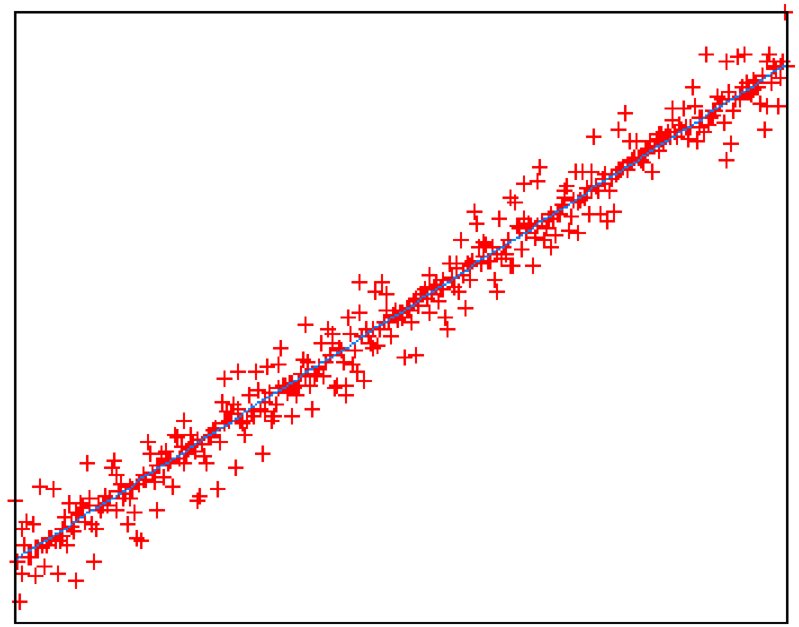
mean square difference between predicted value and actual measurement

Applications: Data fitting 1

Analysis:

linearity:	*f(x)* is nonlinear
convexity:	?? (probably yes)
constrained:	no
smooth:	yes
derivatives:	available, and easy to compute in practice
continuous:	yes, not discrete
ODE/PDE:	no, algebraic

Applications: Data fitting 2



Mathematical description:

$x=\{a,b\}$ parameters for the model $y(t)=at+b$

$f(x)=1/N \sum_i |y_i-y(t_i)|^2$

mean square difference between predicted value and actual measurement

→ *least squares* problem

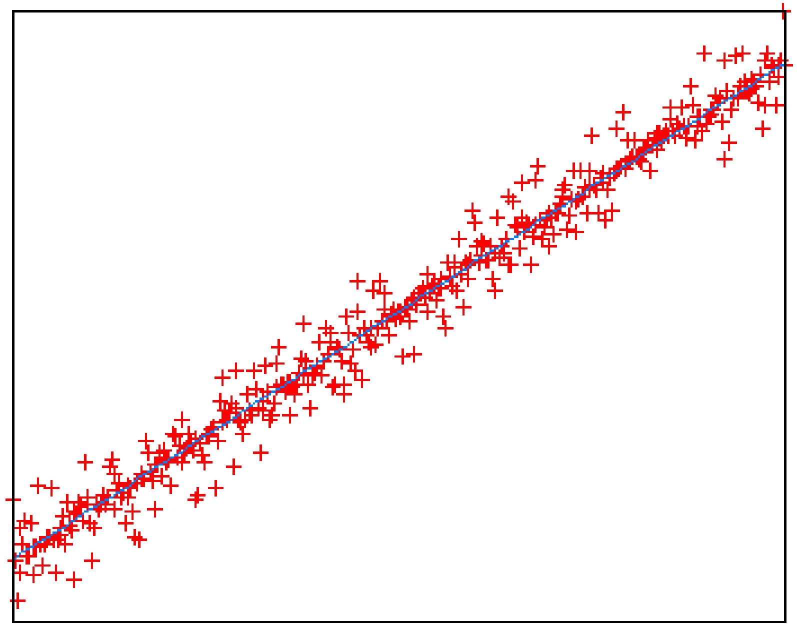
Applications: Data fitting 2

Analysis:

linearity:	*f(x)* is quadratic
Convexity:	yes
constrained:	no
smooth:	yes
derivatives:	available, and easy to compute in practice
continuous:	yes, not discrete
ODE/PDE:	no, algebraic

Note: Quadratic optimization problems (even with linear constraints) are easy to solve!

Applications: Data fitting 3



Mathematical description:

x={a,b} parameters for the model $y(t) = at + b$

$f(x) = 1/N \sum_i |y_i - y(t_i)|$

mean *absolute* difference between predicted value and actual measurement

→ *least absolute error* problem

Applications: Data fitting 3

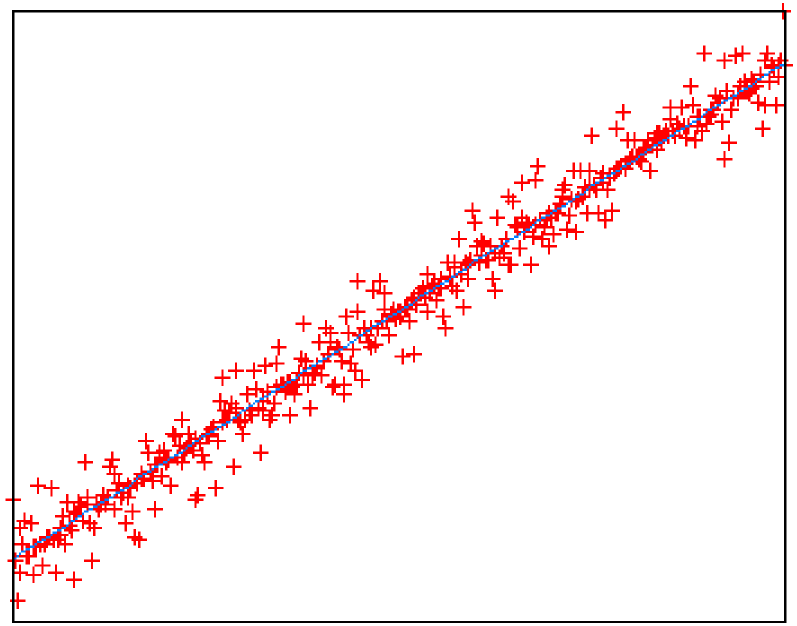
Analysis:

linearity:	*f(x)* is nonlinear
Convexity:	yes
constrained:	no
smooth:	no!
derivatives:	not differentiable
continuous:	yes, not discrete
ODE/PDE:	no, algebraic

Note: Non-smooth problems are really hard to solve!

Wolfgang Bangerth

Applications: Data fitting 3, revisited



Mathematical description:

$x=\{a,b, s_i\}$ parameters for the model $y(t)=at+b$

"slack" variables s_i

$f(x)=1/N \sum_i s_i \rightarrow$ min!

$s_i - |y_i-y(t_i)| \geq 0$

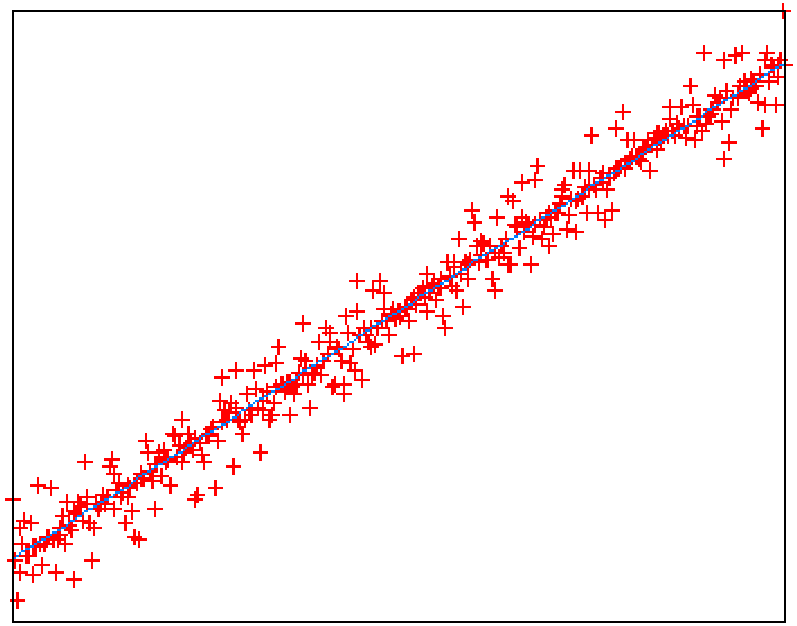
Applications: Data fitting 3, revisited

Analysis:

linearity:	*f(x)* is linear, *h(x)* is not linear
Convexity:	yes
constrained:	yes
smooth:	no!
derivatives:	not differentiable
continuous:	yes, not discrete
ODE/PDE:	no, algebraic

Note: Non-smooth problems are really hard to solve!

Applications: Data fitting 3, re-revisited



Mathematical description:

$x=\{a,b,s_i\}$ parameters for the model $y(t)=at+b$

"slack" variables s_i

$f(x)=1/N \sum_i s_i \rightarrow$ min!

~~$s_i - |y_i - y(t_i)| \geq 0$~~

$s_i - (y_i - y(t_i)) \geq 0$

$s_i + (y_i - y(t_i)) \geq 0$

Applications: Data fitting 3, re-revisited

Analysis:

linearity:	*f(x)* is linear, *h(x)* is now also linear
Convexity:	yes
constrained:	yes
smooth:	yes
derivatives:	yes
continuous:	yes, not discrete
ODE/PDE:	no, algebraic

Note: Linear problems with linear constraints are simple to solve!

Applications: Traveling salesman

Task: Find the shortest tour through *N* cities with mutual distances d_{ij}.

(Here: the 15 biggest cities of Germany; there are 43,589,145,600 possible tours through all these cities.)

Mathematical description:

$x=\{c_i\}$ the index of the *i*th city on our trip, *i=1...N*

$f(x)=\sum_i d_{c_i c_{i+1}}$

$c_i \neq c_j$ for $i \neq j$ no city is visited twice (alternatively: $c_i c_j \geq 1$)

Applications: Traveling salesman

Analysis:

linearity: *f(x)* is linear, *h(x)* is nonlinear

Convexity: meaningless

constrained: yes

smooth: meaningless

derivatives: meaningless

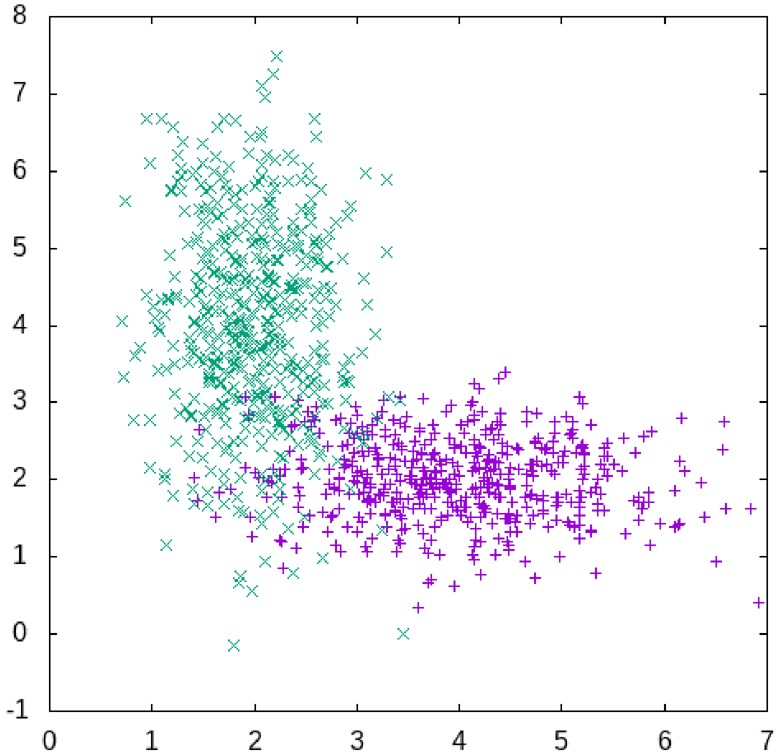
continuous: discrete: $x \in X \subset \{1,2,\ldots,N\}^N$

ODE/PDE: no, algebraic

Note: Integer problems (combinatorial problems) are often exceedingly complicated to solve!

Wolfgang Bangerth

Applications: Classification problems



Task: Find a line that as best as possible separates the two known data sets.

Goal: When a new point comes in, be able to *classify* it with high probability as either green or purple.

Challenge: This often happens in very high dimensions.

Mathematical description:

$x=\{a,b\}$ Coefficients of the line $y=ax+b$

$f(x)$ Number of misclassified green/purple points

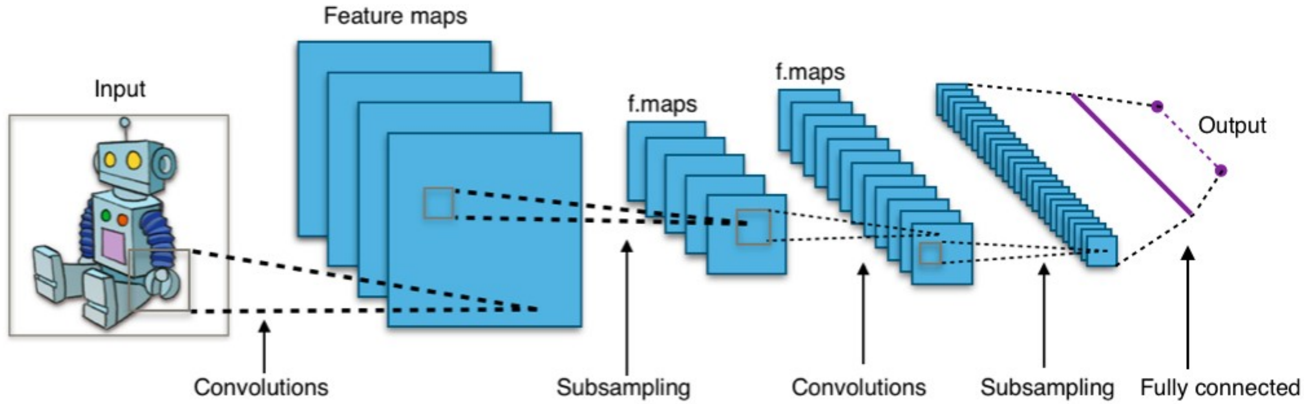
Applications: Classification problems

Analysis:

linearity:	*f(x)* is nonlinear – in fact, it is a step function!
Convexity:	no
constrained:	no
smooth:	no
derivatives:	no (step function)
continuous:	yes, not discrete
ODE/PDE:	no, algebraic

Note: Non-smooth problems are difficult to solve. We may do well reformulating the problem to something smooth.

Applications: Neural networks



Method: Each layer of the NN can be thought of as a parameterized function with inputs from the previous layer.

Task: Find parameters so that we get desired outputs for known inputs.

Mathematical description:

x^n_{ij}	Parameters of node *i,j* of layer *n*
f(x)	Average difference between desired and obtained classification

Applications: Neural networks

Analysis:

linearity:	*f(x)* is nonlinear
Convexity:	?
constrained:	maybe (e.g. if weights have to be positive)
smooth:	yes
derivatives:	depends on formulation
continuous:	yes, not discrete
ODE/PDE:	no, algebraic

Part 2

Minima, minimizers, sufficient and necessary conditions