# Optimization and Modeling

Michael C. Ferris

Computer Sciences Department and Wisconsin Institute for Discovery
University of Wisconsin-Madison

Madison, Wisconsin
October 25, 2023

- Data science: motivation (slides adapted from Steve Wright)
- Optimization: basic components and tradeoffs
- Modeling: using constraints to add domain knowledge
- Uncertainty: how to deal with randomness

# Data Science

- Extract meaning from data: learning
- Use this knowledge to make predictions: inference
- Optimization provides tools for modeling / formulation / algorithms
- Modeling and domain-specific knowledge is vital in practice: "80% of data analysis is spent on the process of cleaning and preparing the data."

# Typical Setup

After cleaning and formatting, obtain a data set of $m$ objects:

- Vectors of features: $a_j$, $j = 1, 2, \ldots, m$
- Outcome / observation / label $y_j$ for each feature vector

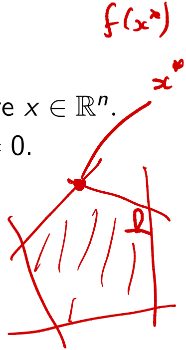
The outcomes $y_j$ could be:

- a real number: regression
- a label indicating the $a_j$ lies in one of $M$ classes (for $M \geq 2$): classification. ($M$ can be very large)
- no labels ($y_j$ is null):
  - subspace identification: locate low-dimensional subspaces that approximately contain the (high-dimensions) vectors $a_j$
  - clustering: partition the $a_j$ into clusters; each cluster groups objects with similar features.

# Optimization

- There is an objective (function) which we are seeking to maximize or minimize described by:

$$f : S \subseteq \mathbb{R}^n \to \mathbb{R} \cup \{+\infty / -\infty\} =: \bar{\mathbb{R}}$$




- Objective function of variables (or unknowns) $f(x)$ where $x \in \mathbb{R}^n$.
- Variables could be subject to constraints such as $h(x) = 0$.
- The feasible set is described by

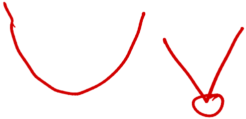
$$\Omega = \{x | h(x) = 0, \ g(x) \leq 0\}$$

- This generates a program of the form

$$\min_x f(x) \text{ s.t. } x \in \Omega$$

- Unconstrained problems have $\Omega = \mathbb{R}^n$ which is the whole space
- What about $\Omega \neq \mathbb{R}^n$?
- Constrained problems can be treated in various ways, including nonlinear, nonconvex problems and convex cones for example.

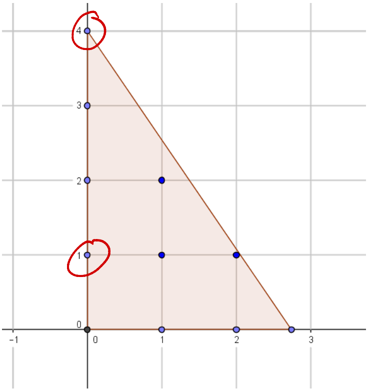
# Four components to optimization



1. Calculus (analysis, probability)
2. Geometry or structure (convexity, polyhedral, discrete)
3. Computation (using linear algebra and sparse tools)
4. Data

- Iterative algorithms generate a series of points which hopefully converge to the solution: issues about well defined (computable), how fast, what they converge to, and how to check properties of the end point.
- Will need all four components; understanding how they link together is important for full command of optimization

# Continuous Vs. Discrete




MIP

- In a discrete problem, only the points would be feasible. In a continuous problem, the whole shaded region is feasible.
- Use case: discrete entities, logic

"$f(x) > 0 \Rightarrow g(x) \le 0$"

# Linear vs. Nonlinear / Stochastic vs Deterministic

- If $f$, $g_i$, $h_i$ are affine then we have a linear program.

$$\begin{array}{lll} \min & f(x) & \\ \text{s.t.} & g_i(x) & \leq 0 \\ & h_i(x) & = 0 \end{array}$$

$g_{\underline{\omega}}$ mean

- Linear problems tend to come from the decision sciences whereas nonlinear problems often arise from physical systems.
- A problem is stochastic if data is not known beforehand. It may arise from some known distribution or assumed via statistical measurements.
- Note the difference between stochastic data and stochastic programs and stochastic algorithms.

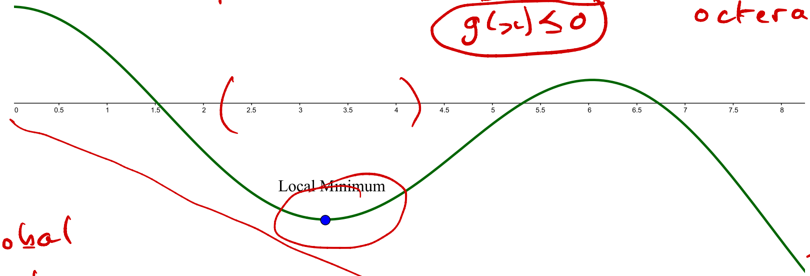
# Global vs Local

We use the notion of local and global minimizers

$\boxed{\text{black box}}$ $f(x)$

→ baron octeract

$g(x) \le 0$




global solvers

$\boxed{\text{minlp}}$

The local minimum is clearly a minimum only within its neighborhood.

Convex functions are ones for which local minimizers are global minimizers.

# Fundamental Data Analysis

Seek a function $\phi$ that:

- approximately maps $a_j$, to $y_j$ for each $j$; $\phi(a_j) \approx y_j$, $j = 1, 2, \ldots, m$
- satisfies additional properties to make it "plausible" for the application, robust to perturbations in the data, generalizable to other data samples from the same distribution.

Can usually define $\phi$ in terms of some parameter vector $x$ - thus identification of $\phi$ becomes a data-fitting problem:

- Find a nice $x$ such that $\phi(a_j; x) \approx y_j$ for $j = 1, 2, \ldots, m$

- Objective function in this problem often built up of $m$ terms that capture mismatch between predictions and observations for data item $(a_j, y_j)$
- The process of finding $\phi$ is called learning or training.

# What's the use of the mapping $\phi$?

- Prediction: Given new data vector $a_k$, predict outputs $y_k \leftarrow \phi(a_k; x)$.
- Analysis: $\phi$ (more particularly the parameter $x$) reveals structure in the data

Many possible complications:

- Noise or errors in $a_j$ and $y_j$
- Missing data:
- Overfitting: $\phi$ exactly fits the set of training data $(a_j, y_j)$ but predicts poorly on "out-of-sample" data $(a_k, y_k)$

# ML models in practice

- Regression: $\phi(a_j; x) = a_j^T x$.

$$\min_x f(x) := \frac{1}{2} \sum_{j=1}^{m} (a_j^T x - y_j)^2$$

- Add $\ell_2 = \|x\|^2$ reduces sensitivity to noise in $y$
- Add $\ell_1 = \|x\|_1$ yields solutions $x$ with few non-zeros (Feature selection)
- loss function $+\lambda * R(x)$
- Sparse PCA.
- Linear Support Vector Machines (kernel SVM)
- Logistic Regression
- Deep learning

All of these modelscan be augmented by domain specific knowledge, leading to nonlinear and/or constrained optimization

# What is meant by a model?

- Many of us build (computer/mathematical) models that capture physics, dynamics, stochastics, discrete choices, and to some extent behavior: collaboration, competition
- Model of system $\boxed{m_x(s, d)}$
- Actions or designs $d$ affect state $s$, parameters $x$
  energy example: state $s =$ electricity flow, actions $d =$ investment/operations, parameters $x =$ loss rate/fuel cost
- Optimization determines model parameters $x$ (based on data - machine learning) (training)
- Can use $m_x(s, d)$ to predict state evolution or specfic outcomes
- Validation ensures predictions are good (testing)

# What does optimization add?

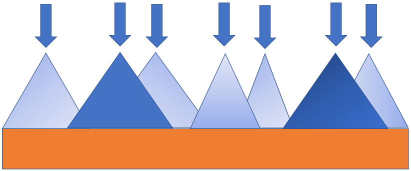
- Value of outcome $v(s, d)$ (e.g $s$ electricity flows in network, $d$ capacity expansion, $v$ is operation profit)
- How to use model to suggest good actions/designs?
- Constrained optimization chooses (feasible) actions to maximize value

$$\max_{s,d} v(s,d) \text{ s.t. } m_x(s,d), (s,d) \in \mathcal{F}$$

- Optimization can be hard to solve (non-convex)
- Models can be complex and difficult to explain, often ignored by decision makers, yet their solution can lead to fundamentally new insights
- Simple rules (policies) $d = \pi(s)$, reduce complexity of optimization, enhance explainability

# Planning models treat uncertainty at different time scales

- Demand growth, technology change, capital costs are long-term uncertainties (years)
- Seasonal inflows to hydroelectric reservoirs are medium-term uncertainties (weeks)
- Levels of wind and solar generation are short-term uncertainties (half hours)
- Very short term effects from random variation in renewables and plant failures (seconds)




- Tradeoff: Uncertainty, cost and operability, regulations, security/robustness/resilience
- Needs modelling at finer time scales

# Simplified two-stage stochastic optimization model

- Investment decisions are $z$ at cost $K(z)$
- Operating decisions are: generation $y$ at cost $C(y)$, loadshedding $q$ at cost $Vq$.
- Random demand is $d(\omega)$.
- Minimize capital cost plus expected operating cost:

Risked

$$
\begin{aligned}
\text{P:} \quad \min_{z,y,q \in X} \quad & K(z) + \mathbb{E}_\omega[C(y(\omega)) + Vq(\omega)] \\
\text{s.t.} \quad & y(\omega) \leq z, \\
& y(\omega) \geq d(\omega) - q(\omega), \\
& z_{\mathcal{N}} \leq (1-\theta) z_{\mathcal{N}}(2017)
\end{aligned}
$$

Who do you have on your bench, what reserves are in your plan?

# Approaches

Value of (constrained) optimization

- Constraints can capture domain knowledge much more than a single objective
- Machine learning can be used to inform models
- Informed strategic decisions and tradeoffs
- Facility location: where to locate reserves, agents, sizing
- Disaster recovery: hedging risk, promoting flexibility, dynamics, windows and staging
- Risk models: not all outcomes are equally bad, trade risk
- Extreme event models: small tails, change policies

Truth is in the details