Lecture 16: Nonlinear Optimization

ISyE 6673: Financial Optimization

Dr. Arthur Delarue
arthur.delarue@isye.gatech.edu

Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

October 26, 2022

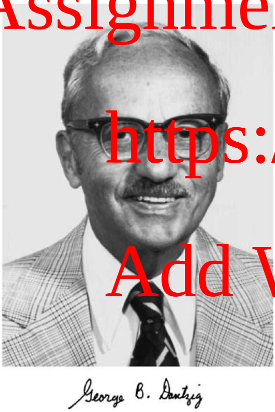
Recap

- Farkas' lemma underpins fundamental theorems in asset pricing
- Markets that obey the law of one price have linear pricing rules
- Complete markets have a unique pricing rule
- Arbitrage-free markets have positive pricing rules

Today

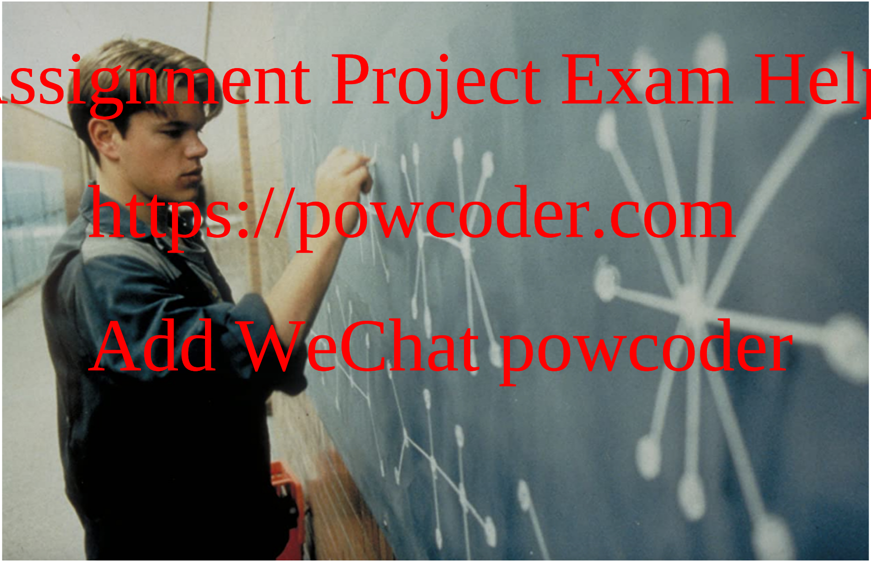
Beyond linear optimization, towards optimization under uncertainty.

History



George Dantzig
inventor of linear programming
and the simplex method

History



History

History

Wisconsin Econometrics Society, 1948

History

Wisconsin Econometrics Society, 1948

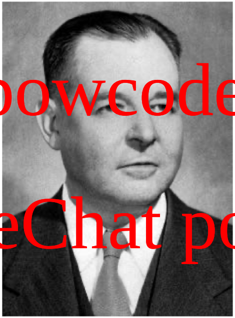
I was a young unknown and I remember how frightened I was with the idea of presenting for the first time to such a distinguished audience the concept of linear programming.

History

Wisconsin Econometrics Society, 1948

History

Wisconsin Econometrics Society, 1948



Harold Hotelling
prominent economist
and statistician

History

Wisconsin Econometrics Society, 1948

History

Wisconsin Econometrics Society, 1948

History

Wisconsin Econometrics Society, 1948

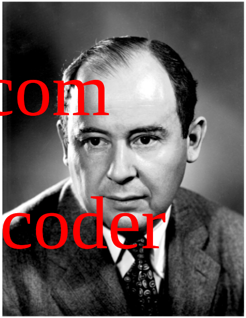
History

Wisconsin Econometrics Society, 1948

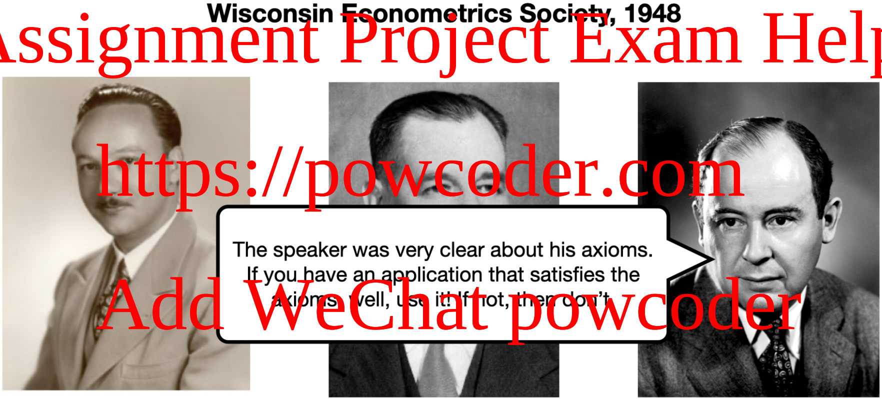
John von Neumann

mathematician, physicist, computer scientist, engineer, economist

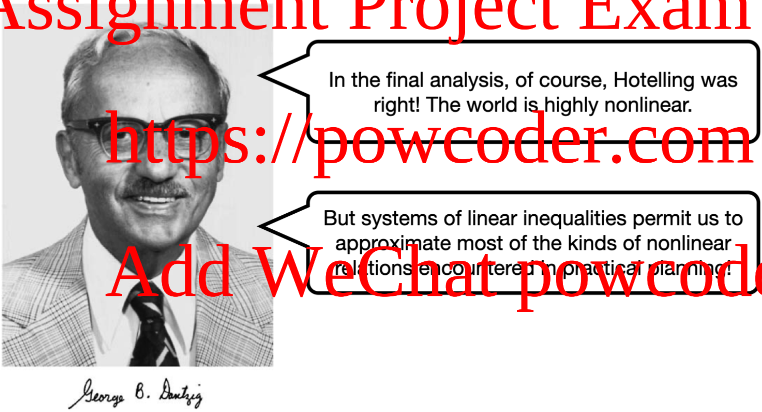
perhaps the most brilliant mind of the 20th century



History


Wisconsin Econometrics Society, 1948
The speaker was very clear about his axioms.
If you have an application that satisfies the
axioms, well, use it! If not, then don't.

History

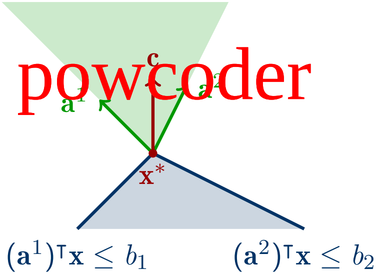

In the final analysis, of course, Hotelling was right! The world is highly nonlinear.
But systems of linear inequalities permit us to approximate most of the kinds of nonlinear relations encountered in practical planning!
George B. Dantzig

A two-dimensional LP

$$(P) \quad \max \quad c_1x_1 + c_2x_2$$
$$\text{s.t.} \quad (\mathbf{a}^i)^\intercal \mathbf{x} \leq b_i \quad 1 \leq i \leq m$$

$$(D) \quad \min \quad \sum_{i=1}^{m} b_i p_i$$
$$\text{s.t.} \quad \sum_{i=1}^{m} p_i \mathbf{a}^i = \mathbf{c}$$
$$\mathbf{p} \geq 0$$

- $\mathbf{x}^*$ primal optimal nondegenerate with active constraints $\mathbf{a}^1$, $\mathbf{a}^2$
- Equivalent to $\mathbf{c} \in \text{cone}(\mathbf{a}^1, \mathbf{a}^2)$



From LP to NLP

- Now consider a two-dimensional nonlinear optimization problem:

$$\begin{aligned} \text{min} \quad & c(\mathbf{x}) \\ \text{s.t.} \quad & f_i(\mathbf{x}) \leq 0 \qquad\qquad 1 \leq i \leq m \end{aligned}$$

- Assume c and f_i are all differentiable convex functions
- Consider an optimal solution $\mathbf{x}^*$ and take the Taylor expansion of all nonlinear functions around $\mathbf{x}^*$:

$$\begin{aligned} \text{min} \quad & c(\mathbf{x}^*) + \nabla c(\mathbf{x}^*)^\intercal(\mathbf{x} - \mathbf{x}^*) \\ \text{s.t.} \quad & f_i(\mathbf{x}^*) + \nabla f_i(\mathbf{x}^*)^\intercal(\mathbf{x} - \mathbf{x}^*) \leq 0 \qquad 1 \leq i \leq m \end{aligned}$$

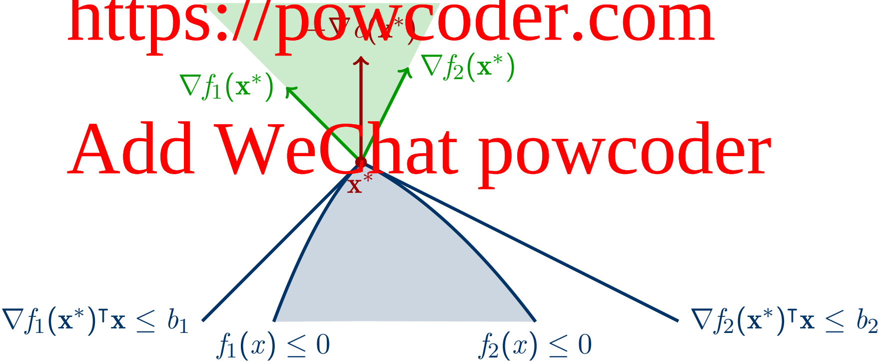
- Equivalently:

$$\begin{aligned} \text{max} \quad & -\nabla c(\mathbf{x}^*)^\intercal \mathbf{x} \\ \text{s.t.} \quad & \nabla f_i(\mathbf{x}^*)^\intercal \mathbf{x} \leq \nabla f_i(\mathbf{x}^*)^\intercal \mathbf{x}^* - f_i(\mathbf{x}^*) \qquad 1 \leq i \leq m \end{aligned}$$

From LP to NLP

$$\begin{aligned} \max \quad & -\nabla c(\mathbf{x}^*)^\intercal \mathbf{x} \\ \text{s.t.} \quad & \nabla f_i(\mathbf{x}^*)^\intercal \mathbf{x} \le \nabla f_i(\mathbf{x}^*)^\intercal \mathbf{x}^* - f_i(\mathbf{x}^*) := b_i \qquad 1 \le i \le m \end{aligned}$$

$\mathbf{x}^*$ optimal, f_1, f_2 active $\Leftrightarrow -\nabla c(\mathbf{x}^*) \in \text{cone}(\nabla f_1(\mathbf{x}^*), \nabla f_2(\mathbf{x}^*))$



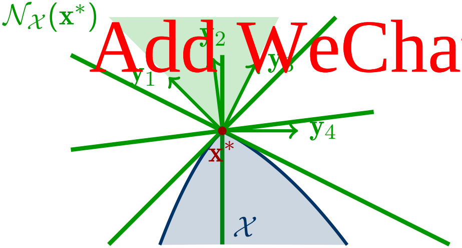
Normal cone

- We can define $\text{cone}(\nabla f_1(\mathbf{x}^*), \nabla f_2(\mathbf{x}^*))$ a bit more generally

Definition

Given a convex set $\mathcal{X}$ and a point $\mathbf{x}^*$ on its boundary, the *normal cone* of $\mathcal{X}$ at $\mathbf{x}^*$ is the set

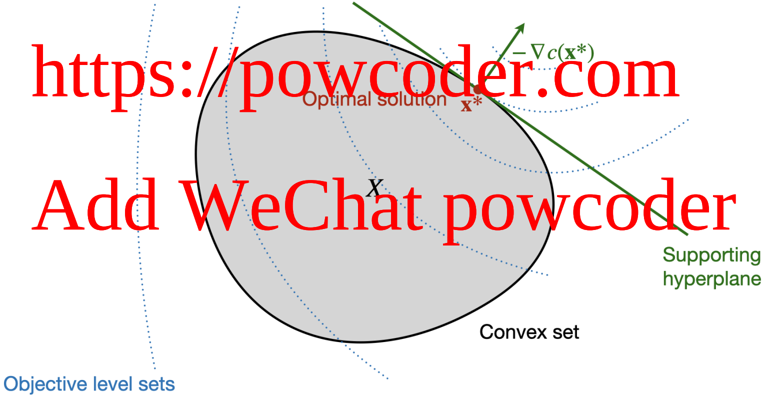
$$\mathcal{N}_{\mathcal{X}}(\mathbf{x}^*) = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y}^\intercal(\mathbf{x} - \mathbf{x}^*) \leq 0 \ \forall \mathbf{x} \in \mathcal{X}\}$$



- $\mathbf{y}^\intercal \mathbf{x} \leq \mathbf{y}^\intercal \mathbf{x}^* \ \forall \mathbf{x} \in \mathcal{X}$
- $\mathbf{y}_1 \in \mathcal{N}_{\mathcal{X}}(\mathbf{x}^*)$
- $\mathbf{y}_2 \in \mathcal{N}_{\mathcal{X}}(\mathbf{x}^*)$
- $\mathbf{y}_3 \in \mathcal{N}_{\mathcal{X}}(\mathbf{x}^*)$
- $\mathbf{y}_4 \notin \mathcal{N}_{\mathcal{X}}(\mathbf{x}^*)$

Supporting hyperplane

- Each vector $\mathbf{y}$ in the normal cone defines a *supporting hyperplane* of the set $\mathcal{X}$ (generalization of a "tangent line")



Geometric optimality conditions

- Convex optimization problem over set $\mathcal{X}$:

$$\begin{aligned} \min \quad & c(\mathbf{x}) \\ \text{s.t.} \quad & f_i(\mathbf{x}) \le 0 \qquad 1 \le i \le m \end{aligned}$$

- A point $\mathbf{x}^*$ is an optimal solution if and only if $\mathbf{x}^*$ is feasible and

$$-\nabla c(\mathbf{x}^*) \in \mathcal{N}_{\mathcal{X}}(\mathbf{x}^*)$$

- Equivalently

$$\nabla c(\mathbf{x}^*)^\intercal(\mathbf{x} - \mathbf{x}^*) \ge 0, \ \forall \mathbf{x} \in \mathcal{X}$$

Algebraic optimality conditions

- A point $\mathbf{x}^*$ is an optimal solution if and only if $\mathbf{x}^*$ is feasible and

$$-\nabla c(\mathbf{x}^*) \in \mathcal{N}_{\mathcal{X}}(\mathbf{x}^*) = \text{cone}\left(\{\nabla f_i(\mathbf{x}^*)\}_{i \text{ active}}\right)$$

- We can write membership in the normal cone algebraically:

$$-\nabla c(\mathbf{x}^*) = \sum_{i=1}^{m} \lambda_i \nabla f_i(\mathbf{x}^*) \qquad \text{(linear combination)}$$

$$\lambda_i \geq 0 \quad \forall i \in [m] \qquad \text{(conic combination)}$$

$$\lambda_i f_i(\mathbf{x}^*) = 0 \quad \forall i \in [m] \qquad \text{(active constraints only)}$$

- These are called the *Karush-Kuhn-Tucker (KKT)* conditions

KKT conditions

KKT conditions for general nonlinear problems

Consider a convex optimization problem (differentiable):

$$
\begin{aligned}
\min \quad & c(\mathbf{x}) & & \text{(P)}\\
\text{s.t.} \quad & f_i(\mathbf{x}) \leq 0 & 1 \leq i \leq m &
\end{aligned}
$$

If $\mathbf{x}^*$ is an optimal solution of (P) and the gradients of the active constraints at $\mathbf{x}^*$ are linearly independent, the following conditions hold:

1. **Stationarity:** there exist multipliers $\lambda_i \geq 0$ for $i \in [m]$ such that

$$\nabla c(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\mathbf{x}^*) = 0$$

2. **Complementary slackness:** for each constraint $i \in [m]$

$$\lambda_i f_i(\mathbf{x}^*) = 0$$

Equality constraints

- Consider the more general convex optimization problem:

$$
\begin{aligned}
\min \quad & c(\mathbf{x}) & & \text{(P)}\\
\text{s.t.} \quad & f_i(\mathbf{x}) \le 0 & 1 \le i \le m &\\
\text{s.t.} \quad & h_k(\mathbf{x}) = 0 & 1 \le k \le r &
\end{aligned}
$$

- Recall $h_k(\mathbf{x}) = 0 \Leftrightarrow h_k(\mathbf{x}) \le 0$ and $-h_k(\mathbf{x}) \le 0$
- Introduce multipliers $\mu_k^+ \ge 0$, $\mu_k^- \ge 0$ and write the stationarity conditions:

$$\nabla c(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_k^+ \nabla h_k(\mathbf{x}^*) - \sum_{i=1}^{m} \mu_k^- \nabla h_k(\mathbf{x}^*) = 0$$

$$\nabla c(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{m} (\mu_k^+ - \mu_k^-) \nabla h_k(\mathbf{x}^*) = 0$$

KKT conditions

KKT conditions for convex problems with equality constraints

If $\mathbf{x}^*$ is an optimal solution and the gradients of the active constraints at $\mathbf{x}^*$ are linearly independent, the following hold:

1. **Stationarity:** there exist multipliers $\lambda_i \geq 0$ for $i \in [m]$ and $\mu_k \in \mathbb{R}$ for $k \in [r]$ such that

$$\nabla c(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_k \nabla h_k(\mathbf{x}^*) = 0$$

2. **Complementary slackness:** for each constraint $i \in [m]$

$$\lambda_i f_i(\mathbf{x}^*) = 0$$

- Multipliers for equality constraints are unconstrained
- Complementary slackness for equality constraints already follows from primal feasibility

Are the KKT conditions ever sufficient?

- The KKT conditions are necessary and sufficient in problems where strong duality holds
- In convex optimization, strong duality doesn't hold automatically, but holds under some *regularity* assumptions

Slater's condition

There exists a feasible $\mathbf{x}$ such that $f_i(\mathbf{x}) < 0$ for all $i \in [m]$.

- If Slater's condition is satisfied, the KKT conditions are both necessary and sufficient!

What about nonconvex problems?

- There exist analogs to the KKT conditions for nonconvex problems, with two caveats

- For nonconvex problems typically the most we can certify is a local optimal solution (not necessarily global)
- If the problem becomes really ugly, the stationarity conditions can become more complex to define

- We can use the KKT conditions for nonlinear nonconvex problems, but we need to be aware of their limitations!

Quadratic programming

- Convex quadratic program (QP):

$$\begin{aligned} \min \quad & \frac{1}{2}\mathbf{x}^\intercal \mathbf{P}\mathbf{x} + \mathbf{q}^\intercal \mathbf{x} + r \\ \text{s.t.} \quad & \mathbf{G}\mathbf{x} \leq \mathbf{h} \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

($\mathbf{P}$ is positive semi-definite)

- Convex quadratically constrained quadratic program (QCQP):

$$\begin{aligned} \min \quad & \frac{1}{2}\mathbf{x}^\intercal \mathbf{P}_0\mathbf{x} + \mathbf{q}_0^\intercal \mathbf{x} + r_0 \\ \text{s.t.} \quad & \frac{1}{2}\mathbf{x}^\intercal \mathbf{P}_i\mathbf{x} + \mathbf{q}_i^\intercal \mathbf{x} + r_i & \qquad 1 \leq i \leq m \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

(all $\mathbf{P}_i$ are positive semi-definite)

Unconstrained quadratic optimization

- Minimizing a quadratic function

$$v^* = \min f_0(\mathbf{x}) := \frac{1}{2}\mathbf{x}^\intercal \mathbf{P}\mathbf{x} + \mathbf{q}^\intercal \mathbf{x} + r$$

- Optimality condition (KKT):

$$\nabla f_0(\mathbf{x}) = \mathbf{P}\mathbf{x} + \mathbf{q} = 0$$

  - If $\mathbf{q} \notin$ span$(\mathbf{P})$, then $v^* = -\infty$
  - If $\mathbf{P}$ is invertible, then $\mathbf{x}^* = -\mathbf{P}^{-1}\mathbf{q}$ is the unique optimum
  - If $\mathbf{P}$ is singular but $\mathbf{q} \in$ span$(\mathbf{P})$, there may be multiple optimal solutions

Least-squares regression

Example

- We are given some data points $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ and want to fit a linear model
- Model parameters are vector of coefficients $\mathbf{w}$ (assume the first feature of $\mathbf{x}$ is constantly 1 so no intercept is needed)
- Optimization problem:

$$\min \sum_{i=1}^{n} (\mathbf{x}_i^\intercal \mathbf{w} - y_i)^2$$

- Define $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\intercal \\ \vdots \\ \mathbf{x}_n^\intercal \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$; then $\mathbf{w} = (\mathbf{X}^\intercal \mathbf{X})^{-1} \mathbf{X}^\intercal \mathbf{y}$

Quadratic optimization with equality constraints

- Equality-constrained convex QP:

$$
\begin{aligned}
\min \quad & \frac{1}{2}\mathbf{x}^\intercal \mathbf{P}\mathbf{x} + \mathbf{q}^\intercal \mathbf{x} + r \\
\text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}
\end{aligned}
$$

- KKT conditions (just stationarity)

$$
\mathbf{P}\mathbf{x} + \mathbf{q} + \sum_{i=1}^{m} \lambda_i \mathbf{a}_i = 0 = \mathbf{P}\mathbf{x} + \mathbf{q} + \mathbf{A}^\intercal \boldsymbol{\lambda}
$$

- Combine with feasibility conditions $\mathbf{A}\mathbf{x} = \mathbf{b}$ to get the following linear system

$$
\begin{bmatrix} \mathbf{P} & \mathbf{A}^\intercal \\ \mathbf{A} & 0 \end{bmatrix}
\begin{bmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{bmatrix}
=
\begin{bmatrix} -\mathbf{q} \\ \mathbf{b} \end{bmatrix}
$$

Quadratic optimization with inequality constraints

- Inequality-constrained convex QP:

$$\begin{aligned} \min \quad & \frac{1}{2}\mathbf{x}^\intercal \mathbf{P}\mathbf{x} + \mathbf{q}^\intercal \mathbf{x} + r \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \end{aligned}$$

- Stationarity conditions:

$$\mathbf{P}\mathbf{x} + \mathbf{q} + \mathbf{A}^\intercal \boldsymbol{\lambda} = 0, \boldsymbol{\lambda} \geq 0$$

- Complementary slackness:

$$\lambda_i(b_i - \mathbf{a}_i^\intercal \mathbf{x}) = 0$$

- Nonlinear!

Nonlinear system of equations

- We can define slack variables $\mathbf{s} = \mathbf{b} - \mathbf{A}\mathbf{x} \geq 0$

$$\lambda_i(b_i - \mathbf{a}_i^\intercal \mathbf{x}) = 0 \Leftrightarrow \lambda_i s_i = 0$$

- Allows to reformulate the KKT conditions as the following nonlinear system:

$$\begin{bmatrix} \mathbf{P}\mathbf{x} + \mathbf{A}^\intercal \boldsymbol{\lambda} + \mathbf{q} \\ \mathbf{A}\mathbf{x} + \mathbf{s} - \mathbf{b} \\ \mathbf{S}\boldsymbol{\Lambda}\mathbf{e} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

where

$$\mathbf{S} = \begin{bmatrix} s_1 & 0 & \dots \\ 0 & s_2 & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}, \ \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots \\ 0 & \lambda_2 & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}, \ \mathbf{e} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \Rightarrow \mathbf{S}\boldsymbol{\Lambda}\mathbf{e} = \begin{bmatrix} \lambda_1 s_1 \\ \vdots \\ \lambda_m s_m \end{bmatrix}$$

Newton's method

Root-finding

Given a differentiable function f, how to find x such that $f(x) = 0$?

Newton's iterative idea

1. Start with a guess x^k
2. Replace f with its first-order Taylor approximation around x^k:

$$f(x^k + d) \approx f(x^k) + f'(x_k) \cdot d = 0$$

3. Solve for the step d

$$f'(x_k) \cdot d = -f(x^k) \Rightarrow d = -\frac{f(x^k)}{f'(x_k)}$$

4. Update $x^{k+1} = x^k + d$ and go to step 2

Newton's method in higher dimensions

- Consider the linear system:

$$f_1(\mathbf{x}) = 0$$
$$f_2(\mathbf{x}) = 0$$
$$f_3(\mathbf{x}) = 0$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$ are all differentiable

- Starting from a guess $\mathbf{x}^k$, we can obtain $\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{d}$ by solving

$$\nabla f_1(\mathbf{x}^k)^\intercal \mathbf{d} = -f_1(\mathbf{x}^k)$$
$$\nabla f_2(\mathbf{x}^k)^\intercal \mathbf{d} = -f_2(\mathbf{x}^k)$$
$$\nabla f_3(\mathbf{x}^k)^\intercal \mathbf{d} = -f_3(\mathbf{x}^k)$$

Newton's method for quadratic programming

- We want to solve the system

$$\begin{bmatrix} \mathbf{P}\mathbf{x} + \mathbf{A}^\intercal \boldsymbol{\lambda} + \mathbf{q} \\ \mathbf{A}\mathbf{x} + \mathbf{s} - \mathbf{b} \\ \mathbf{S}\boldsymbol{\Lambda}\mathbf{e} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

- Three functions on the left; vector-valued, but it doesn't matter
- Start with a guess $\begin{bmatrix} \mathbf{x}^k \\ \mathbf{s}^k \\ \boldsymbol{\lambda}^k \end{bmatrix}$ with $s^k, \lambda^k > 0$, find $\mathbf{d} = \begin{bmatrix} \mathbf{d}_x \\ \mathbf{d}^s \\ \mathbf{d}_\lambda \end{bmatrix}$ such that

$$\begin{bmatrix} \mathbf{P} & 0 & \mathbf{A}^\intercal \\ \mathbf{A} & \mathbf{I} & 0 \\ 0 & \boldsymbol{\Lambda}^k & \mathbf{S}^k \end{bmatrix} \begin{bmatrix} \mathbf{d}_x \\ \mathbf{d}^s \\ \mathbf{d}_\lambda \end{bmatrix} = \begin{bmatrix} -\mathbf{P}\mathbf{x}^k - \mathbf{A}^\intercal \boldsymbol{\lambda}^k - \mathbf{q} \\ -\mathbf{A}\mathbf{x}^k - \mathbf{s}^k + \mathbf{b} \\ -\mathbf{S}^k \boldsymbol{\Lambda}^k \mathbf{e} \end{bmatrix}$$

Interior-point method

- The direction $\mathbf{d}$ is rescaled so that $\mathbf{s}$ and $\boldsymbol{\lambda}$ always remain strictly positive
- At each iteration the solution is strictly feasible
- This is called an *interior-point algorithm* because we only work with interior points, not boundary points
- Results also work in linear programming; philosophically, this is the opposite of what the simplex method does
- There are many more interior-point algorithms and they are the core of convex optimization

A new perspective on inequalities

- Consider the vector inequality $\mathbf{x} \geq \mathbf{y} \Leftrightarrow x_i \geq y_i \; \forall i$
- Vector inequalities define a *partial ordering* on $\mathbb{R}^n$

Definition

The relation $\succeq$ is a partial ordering on $\mathbb{R}^n$ if it satisfies

- $\mathbf{x} \succeq \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$
- If $\mathbf{x} \succeq \mathbf{y}$ and $\mathbf{y} \succeq \mathbf{x}$ then $\mathbf{x} = \mathbf{y}$
- If $\mathbf{x} \succeq \mathbf{y}$ and $\mathbf{y} \succeq \mathbf{z}$ then $\mathbf{x} \succeq \mathbf{z}$

- Trivial to see that linear inequalities satisfy these properties
- Notice that $\mathbf{x} \geq \mathbf{y} \Leftrightarrow \mathbf{x} - \mathbf{y} \geq 0 \Leftrightarrow \mathbf{x} - \mathbf{y} \in \mathbb{R}^n_+$
- What kind of set is $\mathbb{R}^n_+$?

Convex cones, revisited

- $\mathbb{R}^n_+$ (non-negative orthant) is a *convex cone*!

Theorem

Any convex cone $\mathcal{K}$ induces a partial ordering on $\mathbb{R}^n$:

$$\mathbf{x} \succeq_K \mathbf{y} \Leftrightarrow \mathbf{x} - \mathbf{y} \in \mathcal{K}$$

Useful convex cones

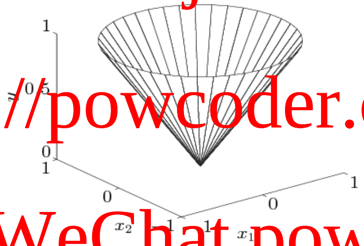
- Non-negative orthant $\mathbb{R}^n_+$
- Norm cones

$$\mathcal{K} = \left\{ \begin{bmatrix} \mathbf{x} \\ t \end{bmatrix} \in \mathbb{R}^n : \|\mathbf{x}\| \le t \right\}$$

  - $\mathcal{K}$ is the epigraph of a (convex) norm function means $\mathcal{K}$ is convex
  - If $(\mathbf{x}, t) \in \mathcal{K}$ then for any $a \ge 0$, $\|a\mathbf{x}\| = a\|\mathbf{x}\| \le at$ so $(a\mathbf{x}, at)$ is in $\mathcal{K}$ (indeed a cone)
- Our favorite norm is the Euclidean norm $\|\cdot\|_2$ which induces a second-order cone:

$$\mathcal{L}^{n+1} = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ t \end{bmatrix} : \sqrt{x_1^2 + \ldots + x_n^2} \le t \right\}$$

Second-order cone



- Boundary of $\mathcal{L}^3 \subseteq \mathbb{R}^3$
- Also called Lorentz cone, quadratic cone, ice-cream cone

Linear conic inequalities

Linear conic inequality (aka generalized inequality)

Given a partial ordering $\succeq_{\mathcal{K}}$ induced by the cone $\mathcal{K}$, we can define a linear conic inequality as

$$\mathbf{Ax} \succeq_{\mathcal{K}} \mathbf{b} \Leftrightarrow \mathbf{Ax} - \mathbf{b} \in \mathcal{K}$$

- Can be nonlinear if $\mathcal{K}$ is nonlinear
- Unlike $\geq$, this is not a component-wise inequality

Second-order inequality example

Example

Is it true that $\begin{bmatrix}1\\2\\3\end{bmatrix} \succeq_{\mathcal{L}^3} \begin{bmatrix}2\\1\\1\end{bmatrix}$?

- $\begin{bmatrix}1\\2\\3\end{bmatrix} - \begin{bmatrix}2\\1\\1\end{bmatrix} = \begin{bmatrix}-1\\1\\2\end{bmatrix}$
- $\sqrt{(-1)^2 + 1^2} = \sqrt{2} \leq 2$

Second-order cone programming (SOCP)

- Second-order cone program (SOCP):

$$\begin{aligned} \min \quad & \mathbf{p}^\top \mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}_i \mathbf{x} \succeq_{\mathcal{L}^{n_i}} \mathbf{b}_i && \forall i \in [m] \\ & \mathbf{G}\mathbf{x} = \mathbf{h} \end{aligned}$$

where $\mathbf{A}_i = \begin{bmatrix} \bar{\mathbf{A}}_i \\ \mathbf{c}_i^\top \end{bmatrix} \in \mathbb{R}^{n_i \times n}$ and $\mathbf{b}_i = \begin{bmatrix} \bar{\mathbf{b}}_i \\ d_i \end{bmatrix} \in \mathbb{R}^{n_i}$

- Equivalently:

$$\begin{aligned} \min \quad & \mathbf{p}^\top \mathbf{x} \\ \text{s.t.} \quad & \left\| \bar{\mathbf{A}}_i \mathbf{x} - \bar{\mathbf{b}}_i \right\|_2 \le \mathbf{c}_i^\top \mathbf{x} - d_i && \forall i \in [m] \\ & \mathbf{G}\mathbf{x} = \mathbf{h} \end{aligned}$$

Recap

- Geometric optimality conditions in LP generalize to NLP as KKT conditions

- Interior-point methods solve convex optimization problems
- Next week: application to portfolio optimization