

CS635: Tools and Environments for Optimization. Lecture 22

Michael Ferris

Computer Sciences Department
University of Wisconsin-Madison

April 18, 2014



Nonlinear Programming: Why?

An anecdote: July, 1948. A young and frightened George Dantzig, presents his newfangled “linear programming” to a meeting of the Econometric Society of Wisconsin, attended by distinguished scientists like Hotelling, Koopmans, and Von Neumann. Following the lecture, Hotelling^a pronounced to the audience:

^ain Dantzig’s words “a huge whale of a man”



Nonlinear Programming: Why?

An anecdote: July, 1948. A young and frightened George Dantzig, presents his newfangled “linear programming” to a meeting of the Econometric Society of Wisconsin, attended by distinguished scientists like Hotelling, Koopmans, and Von Neumann. Following the lecture, Hotelling^a pronounced to the audience:

But we all know the world is nonlinear!

^ain Dantzig’s words “a huge whale of a man”



Nonlinear Programming: Why?

An anecdote: July, 1948. A young and frightened George Dantzig, presents his newfangled “linear programming” to a meeting of the Econometric Society of Wisconsin, attended by distinguished scientists like Hotelling, Koopmans, and Von Neumann. Following the lecture, Hotelling^a pronounced to the audience:

But we all know the world is nonlinear!

^ain Dantzig’s words “a huge whale of a man”

The world is indeed
nonlinear



Nonlinear Programming: Why?

An anecdote: July, 1948. A young and frightened George Dantzig, presents his newfangled “linear programming” to a meeting of the Econometric Society of Wisconsin, attended by distinguished scientists like Hotelling, Koopmans, and Von Neumann. Following the lecture, Hotelling^a pronounced to the audience:

But we all know the world is nonlinear!

^ain Dantzig’s words “a huge whale of a man”

The world is indeed
nonlinear

- Physical Processes and Properties
 - Equilibrium
 - Enthalpy



Nonlinear Programming: Why?

An anecdote: July, 1948. A young and frightened George Dantzig, presents his newfangled “linear programming” to a meeting of the Econometric Society of Wisconsin, attended by distinguished scientists like Hotelling, Koopmans, and Von Neumann. Following the lecture, Hotelling^a pronounced to the audience:

But we all know the world is nonlinear!

^ain Dantzig’s words “a huge whale of a man”

The world is indeed nonlinear

- Physical Processes and Properties
 - Equilibrium
 - Enthalpy
- Abstract Measures
 - Economies of Scale
 - Covariance
 - Utility of decisions



Making Tires

- Tires are made by combining rubber, oil, and carbon.
- Tires must have a “hardness” of between 25 and 35.
- Tires must have an “elasticity” of at least 16
- Tires must have a “tensile strength” of at least 12
- To make a set of four tires, we require 100 pounds of total product (rubber, oil, and carbon).
 - At least 50 pounds of carbon
 - Between 25 and 60 pounds of rubber



Real Engineering

- Chemical Engineers tell you that the tensile strength, elasticity, and hardness of tires made of r pounds of rubber, o pounds of oil, and c pounds of carbon are...
 - Tensile Strength = $12.5 - 0.1o - 0.001o^2$
 - Elasticity = $17 + .35r - 0.04o - 0.002r^2$
 - Hardness =
 $34 + 0.1r + 0.06o - 0.3c + 0.01ro + 0.005o^2 + 0.001c^{1.95}$
- The cost of rubber is \$.04/pound, oil costs \$.01 per pound, and carbon costs \$.07/pound



Nonlinear Optimization Model

$$\min 0.04r + 0.01o + 0.07c$$

$$r \in [25, 60]$$

$$c \geq 50$$

$$o \geq 0$$

$$t \geq 12$$

$$e \geq 16$$

$$h \in [25, 35]$$

$$r + o + c = 100$$

$$12.5 - 0.1o - 0.001o^2 = t$$

$$17 + .35r - 0.04o - 0.002r^2 = e$$

$$34 + 0.1r + 0.06o - 0.3c + 0.01ro +$$

$$0.005o^2 + 0.001c^{1.95} = h$$



Important Safety Tips

- a^x is evaluated as $e^{\log(a^x)} = e^{x \log(a)}$
- This is **bad**¹ when a is negative
- If x is an integer, a^x should **not** be bad
- Use `power(a, x)`. x should be an integer, though.
- There is also a `sqr` function

¹complex



Smallest Sphere

\$1 Problem

- Given points $\{p_1, p_2, \dots, p_m\}$, with $p_j \in \mathbb{R}^n$
- Find the smallest n -dimensional sphere that encloses all the points.



Smallest Sphere

\$1 Problem

- Given points $\{p_1, p_2, \dots, p_m\}$, with $p_j \in \mathbb{R}^n$
 - Find the smallest n -dimensional sphere that encloses all the points.
-
- Set variable levels **before** solving to give good starting point to NLP solvers



Make Some Money in the Market

- R — A random variable representing portfolio return
- a_i — The (random) return of stock i
- w_i — A decision variable. The amount invested in stock i



Make Some Money in the Market

- R — A random variable representing portfolio return
- a_i — The (random) return of stock i
- w_i — A decision variable. The amount invested in stock i

$$R = \sum_{i=1}^N a_i w_i$$



Make Some Money in the Market

- R — A random variable representing portfolio return
- a_i — The (random) return of stock i
- w_i — A decision variable. The amount invested in stock i

$$R = \sum_{i=1}^N a_i w_i$$

- We can't optimize R (since it is random), but we can optimize $\mathbb{E}R$



Expectation is Linear

$$\begin{aligned}\mathbb{E}R &= \mathbb{E} \left[\sum_{i=1}^N a_i w_i \right] \\ &= \sum_{i=1}^N \mathbb{E}[a_i w_i] \\ &= \sum_{i=1}^n w_i \mathbb{E}a_i \stackrel{\text{def}}{=} \sum_{i=1}^n \alpha_i w_i\end{aligned}$$



Optimization Model

$$\max \sum_{i=1}^N \alpha_i w_i$$

subject to

$$\sum_{i=1}^N w_i = B$$
$$w_i \geq 0$$



Optimization Model

$$\max \sum_{i=1}^N \alpha_i w_i$$

subject to

$$\sum_{i=1}^N w_i = B$$
$$w_i \geq 0$$

One Jillion \$ Question

- What is the optimal solution?



Risk

Which of these two yearly returns would you rather have?

Year	Portfolio #1	Portfolio #2
1	100	100
2	100	200
3	100	0
4	100	10
5	100	200
Avg	100	102

- Let's measure **deviation from average**
- Portfolio 1: 0
- Portfolio 2:

$$((102-100)^2 + (102-200)^2 + (102-0)^2 + (102-10)^2 + (102-200)^2)$$

- Variance** is like average deviation from average



Risk

Useful Formulae?

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2) - (\mathbb{E}X)^2 \\ \text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]\end{aligned}$$



More Ugly Math

$$\text{Var}(R) = \text{Var}\left(\sum_{i=1}^N a_i w_i\right)$$



More Ugly Math

$$\begin{aligned}
 \text{Var}(R) &= \text{Var}\left(\sum_{i=1}^N a_i w_i\right) \\
 &= \sum_{i=1}^N \text{Var}(a_i w_i) + 2 \sum_{i=1}^n \sum_{j=i+1}^n \text{Cov}(a_i w_i, a_j w_j) \\
 &= \sum_{i=1}^N w_i^2 \text{Var}(a_i) + 2 \sum_{i=1}^n \sum_{j=i+1}^n w_i w_j \text{Cov}(a_i, a_j) \\
 &= w^T Q w
 \end{aligned}$$

- Where Q is the **variance-covariance** matrix



Optimization

Ingredients for a Model

$$\mathbb{E}(R) = \sum_{i=1}^N \alpha_i w_i$$

$$\text{Var}(R) = \sum_{i=1}^N \sigma_i^2 w_i^2 + 2 \sum_{i=1}^n \sum_{j=i+1}^n w_i w_j \text{Cov}(a_i, a_j) = w^T Q w$$



Optimization

Ingredients for a Model

$$\mathbb{E}(R) = \sum_{i=1}^N \alpha_i w_i$$

$$\text{Var}(R) = \sum_{i=1}^N \sigma_i^2 w_i^2 + 2 \sum_{i=1}^n \sum_{j=i+1}^n w_i w_j \text{Cov}(a_i, a_j) = w^T Q w$$

- How do we put them together?



Models

$$\max \sum_{i=1}^N \alpha_i w_i - \lambda w^T Q w$$

subject to

$$\begin{aligned} \sum_{i=1}^N w_i &= B \\ w_i &\geq 0 \end{aligned}$$

$$\min w^T Q w$$

subject to

$$\begin{aligned} \sum_{i=1}^N \alpha_i w_i &\geq K \\ \sum_{i=1}^N w_i &= B \\ w_i &\geq 0 \end{aligned}$$

$$w^T Q w = \sum_{i=1}^N \sigma_i^2 w_i^2 + 2 \sum_{i=1}^n \sum_{j=i+1}^n w_i w_j \text{Cov}(a_i, a_j)$$



Some Real Data

- `qdata.inc`
- `qp1.gms`
- Note how mean and covariance are computed...



Benchmarks

- b_i : Some numbers such that $\sum_{i \in N} b_i = 1$
- Market-cap based benchmarks:

$$b_i = \frac{v_i}{\sum_{j \in N} v_j}$$

where v_i is market-cap (value) of company i

- S&P 500, Russell 3000, etc.
- Note that benchmark weights b_i change every day
- We would like to “track” a benchmark “closely”
 - Without having to trade too much
 - Without owning too many stocks



Real Portfolio Optimization Models

- Factor models
 - Short Sales
 - Benchmark tracking
 - Limit names
 - Different risk measures
 - Limit transactions
 - Market impact
-



Real Portfolio Optimization Models

- Factor models
- Short Sales
- Benchmark tracking
- Limit names
- Different risk measures
- Limit transactions
- Market impact

MINLP

- Many of these require **MINLP**



General (Nonlinear) Optimization Model

$$\max_{x \in \mathbb{R}^n} f(x)$$

- x is an n -dimensional *vector*:
 $x = (x_1, x_2, \dots, x_n)$.
- $f(x)$: **Objective Function**



General (Nonlinear) Optimization Model

$$\max_{x \in \mathbb{R}^n} f(x)$$

subject to

$$g_i(x) \left\{ \begin{array}{l} \leq \\ = \\ \geq \end{array} \right\} b_i \quad \forall i \in M$$

- x is an n -dimensional *vector*:
 $x = (x_1, x_2, \dots, x_n)$.
- $f(x)$: **Objective Function**
- M : **Constraint Set**
- $g_i(x) \{ \leq, =, \geq \} b_i$: **Constraint**



General (Nonlinear) Optimization Model

$$\max_{x \in \mathbb{R}^n} f(x)$$

subject to

$$g_i(x) \left\{ \begin{array}{l} \leq \\ = \\ \geq \end{array} \right\} b_i \quad \forall i \in M$$

$$x \in X$$

- x is an n -dimensional *vector*:
 $x = (x_1, x_2, \dots, x_n)$.
- $f(x)$: **Objective Function**
- M : **Constraint Set**
- $g_i(x) \{ \leq, =, \geq \} b_i$: **Constraint**
- X : **Explicit Constraint Set** Maybe
 $X \subseteq \mathbb{Z}^n$



Objective Functions

- A *linear function* $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a weighted sum:
 - $f(x_1, \dots, x_n) = \sum_{i=1}^n c_i x_i$ for given coefficients $\{c_1, \dots, c_n\}$.
 - $f(x) = c^T x$



Objective Functions

- A *linear function* $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a weighted sum:
 - $f(x_1, \dots, x_n) = \sum_{i=1}^n c_i x_i$ for given coefficients $\{c_1, \dots, c_n\}$.
 - $f(x) = c^T x$
- Everything that is not linear is nonlinear
 - $f(x) = \sin(x)$
 - $f(x, y, z) = 2x - 3y + 14z^2$
 - $f(x) = x^T Q x = \sum_{i=1}^n \sum_{j=1}^n q_{ij} x_i x_j$
 - $f(x_1, x_2) = (3x_1 - 2x_2 + 7)/(x_1 + x_2)$.



You thought you'd never use Calculus again...

Fundamental Concept

It is *extremely* important to understand the convexity properties of a function you are trying to optimize.



You thought you'd never use Calculus again...

Fundamental Concept

It is *extremely* important to understand the convexity properties of a function you are trying to optimize.

- A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is *convex* if for any two points x and y , the graph of f lies below or on the straight line connecting $(x, f(x))$ to $(y, f(y))$ in \mathbb{R}^{n+1} .
 - $f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) \quad \forall 0 \leq \alpha \leq 1$



Concave = -Convex

- A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is *concave* if for any two points x and y , the graph of f lies above or on the straight line connecting $(x, f(x))$ to $(y, f(y))$ in \mathbb{R}^{n+1} .
 - $f(\alpha x + (1 - \alpha)y) \geq \alpha f(x) + (1 - \alpha)f(y) \quad \forall 0 \leq \alpha \leq 1$
- A function f is convex if and only if the function $-f$ is concave



Concave = -Convex

- A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is *concave* if for any two points x and y , the graph of f lies above or on the straight line connecting $(x, f(x))$ to $(y, f(y))$ in \mathbb{R}^{n+1} .
 - $f(\alpha x + (1 - \alpha)y) \geq \alpha f(x) + (1 - \alpha)f(y) \quad \forall 0 \leq \alpha \leq 1$
- A function f is convex if and only if the function $-f$ is concave
- A function that is neither convex nor concave, we will call *nonconvex*.



Function Pictures

CONVEX

x2

NONCONVEX

x3



Without Drawing Pictures?

- I like drawing pictures nearly as much as my 10-year old son, but we may need a better way to tell if a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex.
- Remember from calculus: A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is **convex** on a domain \mathcal{D} if and only if $f''(x) > 0 \forall x \in \mathcal{D}$
 - Think: x^2 . $f''(x) = 2 > 0$
- The analogy in multiple dimensions, is a square matrix of second partial derivatives called the **Hessian** matrix, denoted $\nabla^2 f(x)$



Gradients and Hessians

- Given a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$

Gradient

- The multi-dimension analog of the **derivative** of a function
- The **gradient** of f , $\nabla(f) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

Hessian

- The multi-dimension analog of the **second derivative** of a function
- The **Hessian** of f , $\nabla^2(f) : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ is a matrix made up of all second partial derivatives



The Hessian

- Written on in full:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Characterization of Convexity

- A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is **convex** on a domain \mathcal{D} if and only if $\nabla^2 f(x)$ is **positive semi-definite** $\forall x \in \mathcal{D}$



Positive What-I-Definite!?!?!?!?

- A matrix A is **positive semi-definite** if and only if any of the following conditions hold:

- 1 $x^T Ax \geq 0 \forall x \in \mathbb{R}^n$
- 2 All eigenvalues of A are ≥ 0



Positive What-I-Definite!?!?!?!?

- A matrix A is **positive-definite** if and only if any of the following conditions hold:

- 1 $x^T Ax > 0 \forall x \neq 0 \in \mathbb{R}^n$
- 2 All eigenvalues of A are > 0
- 3 The determinant of all **leading principal minors** of A are > 0



Easy/Hard?

Wal-Mart

- Problems are “easy” if locally optimal solutions are also globally optimal
-
- Which of the following do you suspect are easy, which are hard?
 - 1 Minimize a convex function?
 - 2 Minimize a concave function?
 - 3 Minimize a nonconvex function?
 - 4 Maximize a convex function?
 - 5 Maximize a concave function?
 - 6 Maximize a nonconvex function?
 - True or False?
 - 1 Linear Functions are Convex?
 - 2 Linear Functions are Concave?



Convexity – Again. Ugh!

- A set S is *convex* if the straight line segment connecting any two points in S lies entirely inside or on the boundary of S .
 - $x, y \in S \Rightarrow \alpha x + (1 - \alpha)y \in S \quad \forall 0 \leq \alpha \leq 1$



Convexity – Again. Ugh!

- A set S is *convex* if the straight line segment connecting any two points in S lies entirely inside or on the boundary of S .

- $x, y \in S \Rightarrow \alpha x + (1 - \alpha)y \in S \quad \forall 0 \leq \alpha \leq 1$

-
- A Confusing Point...

- Why do they have a *convex function* and a *convex set*? How are they related?
- f is convex if and only if the *epigraph*, or “over part” of f is a convex set.



Breaking down the convex sets

- Polyhedral. A Set S is polyhedral if...
 - Formal Definition: it is the intersection of a finite number of half-spaces.
 - Informal Definition: it is “linear”.
 - A good working definition: If the explicit constraint set (X) is empty and all of the $g_i(x)$ are linear functions.
 - $S = \{x \in \mathbb{R}^n : Ax \leq b, x \geq 0\}$ is polyhedral.
- Non-polyhedral
 - Anything “curvy”
 - Anything “discrete”



A 'Canonical' Nonlinear Program (NLP)

$$\max f(x)$$

s.t.

$$g_i(x) \leq b_i \quad \forall i = 1, \dots, m$$

$$x_j \geq 0 \quad \forall j = 1, \dots, n$$

- Problem is “easy” if $f(\cdot)$ is **concave** and $g_i(\cdot)$ is **convex** $\forall i$
- $f(\cdot)$ concave \Rightarrow Local maximum is global maximum
- $g_i(\cdot)$ is **convex** $\forall i \Rightarrow$ the feasible region is a **convex set**



Important Point

- These previous conditions are only **sufficient** conditions for convexity, not necessary

$X_1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$ is convex

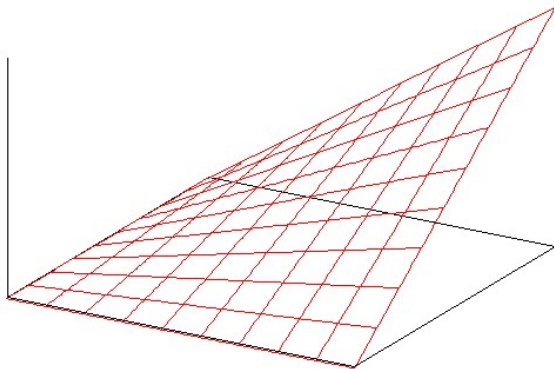
$X_2 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \geq 1\}$ is **not** convex

$X_3 = \{(x, y) \in \mathbb{R}_+^2 \mid xy \geq 1\}$ is convex



Worth 1000 Words?

$$x_i x_j$$



GAMS NLP Solvers

Local Solvers

- CONOPT: Generalized Reduced Gradient
- IPOPT: Interior Point
- KNITRO: Many, but basically Interior Point
- MINOS: Reduce Gradient/Projected Lagrangian
- MOSEK: Interior Point
- PATHNLP: NLP Solver (Not recommended)
- SNOPT: Sequential Quadratic Programming

Maybe Global Solvers

- LGO: No guarantee of optimality? Lipschitz?
- MSNLP: Multi-start method
- OQNLP: Multi-start method



CONVEX Quadratic: QCP

- Add CPLEX, Gurobi, Mosek, Xpress, SCIP²
- These (commercial) solvers should be the fastest. Most also have *MIQCP* capabilities

²SCIP is very good, but not commercial



KKT

- Local Solvers find only **KKT point**: A point satisfying **necessary conditions** for a solution to be optimal
 - And in fact these KKT conditions are only *really* necessary if a **constraint qualification** holds
-



KKT

- Local Solvers find only **KKT point**: A point satisfying **necessary conditions** for a solution to be optimal
- And in fact these KKT conditions are only *really* necessary if a **constraint qualification** holds

Global Solvers

- BARON
- COUENNE
- LINDOGLOBAL



Functions

- multiplication (*), division (/), exponentiation (**)
- abs, arctan, ceil, cos, errorf, exp, floor, log, log10, max, min, mod, normal, power, round, sign, sin, sqr, sqrt, trunc, uniform
- Note that abs, ceil, floor, max, min, mod, normal, round, sign, trunc, uniform cannot be used in NLP models (not differentiable or deterministic)
- log only defined on $x > 0$. sqrt defined on $x \geq 0$, but its derivative is not defined when $x = 0$.



My Model Doesn't Solve



- Try different solvers, CONOPT, KNITRO
- Starting points, maybe use loop to set level (.1)
- Update bounds on variables to restrict search.
- Try using a global solver on a smaller instance



SuperModel – Revenue Management?

- Given I : A set of fare classes
- α_i : % of plane that a seat of fare class $i \in I$ takes up
- c_i : The cost to purchase a ticket in fare class $i \in I$
- d_i : The number of people willing to purchase a ticket in fare class $i \in I$
- ℓ_i : A lower bound on the percentage of the plane dedicated to fare class $i \in I$
- n_i : Percentage of people for fare class $i \in I$ that do not show up
- p_{ij} : Price to move a ticketed passenger from fare class $i \in I$ to $j \in I$



Piecewise Linear

Let's do some “global optimization” by doing piecewise-linear programming.

$$f_i(x) = \alpha_i x^4 + \beta_i x^3 + \gamma_i x^2 + \delta_i x$$

$$g_i(x) = A_i x^3 + B_i x^2 C_i x + D_i$$



A Problem

$$\begin{aligned}
 \max \quad & \sum_{i=1}^n f_i(x_i) \\
 & \sum_{i:i \text{ is even}} g_i(x_i) \geq -3 \\
 & \sum_{i:i \text{ is even}} g_i(x_i) \leq -3 \\
 & \sum_{i:i \text{ is odd}} g_i(x_i) \geq -3 \\
 & \sum_{i:i \text{ is odd}} g_i(x_i) \leq -3 \\
 & -1 \leq x_i \leq 1
 \end{aligned}$$

