Optimization

A (mathematical) optimization problem:

$$\begin{aligned} &\text{minimize} && f_0(\vec{x}) \\ &\text{subject to} && \underbrace{f_i(\vec{x}) \leq b_i}_{\text{'budget'}} , \quad i=1,\dots,m \end{aligned}$$

where

$\vec{x} = (x_1, x_2, \cdots, x_n)$: optimization variables

$f_0 : \mathbb{R}^n \to \mathbb{R}$: objective function

$f_i : \mathbb{R}^n \to \mathbb{R}$, $i=1,\dots,m$: constraint functions

the solution $\vec{x}^*$ to the optimization problem has the smallest value of f_0 among all vectors that satisfy the constraints.

Example: manufactoring

variables: ~~profit,~~ transport, device specs, ad budget

constraints: ad budget,

objective: profit,

Example: portfolio optimization

variables: amounts invested in different assets

constraints: budget, max/min investment per asset, min return

objective: overall risk or return variance

Example: data fitting

variables: model parameters

constraints: prior information, parameter limits

objective: measure of misfit / implausibility, or prediction error

Example 1: PC profit

A PC manufacturer currently sells 10,000 units/ month of a certain model. The manufacturing cost is \$700/ unit. The wholesale price is \$950/ unit. In a few test markets the price was discounted by \$100/ unit, which resulted in a 50% sales increase. The current advertising costs are \$50,000/ month. Research by an ad agency shows an increase of advertising budget by \$10,000 would increase sales by 200 units/month. (The company will not increase the ad budget above \$100,000). Determine the price discount that will maximize profit.

Variables N = total # units sold, q = profit/unit, P = profit

n = current sales = 10,000 units/month

m = manufactoring cost = \$700/unit

w = unit price (wholesale) = \$950/unit

D = % of test discount

d = amount of test discount = \$100/unit

r = % sales increase from test = 50%

a = advertising = \$50,000/month

b = test ad budget increase = \$10,000, B = % of test ad increase

c = increase in sales due to ad test = 200/month

A = cap on ad budget = \$100,000/month

Goal: Maximize P, the total profit

Assumptions:

- Budget can only change for ads + discounts
- No other costs to consider (e.g. transport)

Model

Case 1: base advertizing; maximize profit based on discount

The total profit is

$$P = \underbrace{Nq}_{\text{profit from units sold}} - a - bB$$

a ← loss from advertizing (base)

bB ← loss from adertizing increase

We need to have a model only as a function of D:

$$N = \underbrace{n(1 + rD)}_{\#\text{ sold by increasing discount}} + \underbrace{cB}_{\#\text{ sold by increasing ads}}$$

$$= 10{,}000(1 + 0.5D) + c(0) = 10{,}000(1 + 0.5D)$$

and

$$q = w - m - dD$$

$$= 950 - 700 - 100D = 250 - 100D$$

$\boxed{\text{Check}}$ Dimensions need to be consistent

$$P = Nq - a$$

$$[\$] = [\text{units}][\$/\text{unit}] - [\$] = [\$] \checkmark$$ (all units per month)

$$N = n(1 + rD) + cB$$

$$[\text{units}] = [\text{units}](1 + [-][-]) + [\text{units}][-] \checkmark$$ (per month)

$$q = p - m - dD$$

$$[\$/\text{unit}] = [\$/\text{unit}] - [\$/\text{unit}] - [\$/\text{unit}][-]$$

Maximize: Rewrite P as a function of D.

$$P = n(1 + rD)(w - m - dD) - a$$

$$= 10{,}000(1 + 0.5D)(250 - 100D) - 50{,}000 \qquad \text{(figure)}$$

Find the critical points:

$$\frac{dP}{dD} = 10{,}000\left[(1+0.5D)(-100) + (250-100D)(0.5)\right] = 0$$

$$\rightarrow D^* = 0.25$$

$$\frac{d^2P}{dD^2} < 0 \Rightarrow \text{Maximum (by the } 2^{nd} \text{ Derivative Test)}$$

($f''(c) > 0 \Rightarrow$ local min (concave up)
$f''(c) < 0 \Rightarrow$ local max (concave down))
for each critical pt c

Interpret: the price has to be discounted by \$25 to maximize the total profit

Note: $P(D^*) = \$2{,}481{,}250$/month

Sensitivity Analysis (1.2)

How might we evaluate the robustness of our model. That is, how likely is it for the result to remain true even though the model may not be entirely accurate?

For example, some of the variables used in the PC example had precise values (e.g., manufactoring cost), but some others did not. The % sales increase, for example, might change depending on the size of the test market. We want to see how sensitive the total profit P is to r

Let x^* be the solution to an optimization problem $\max_x f(x)$. Perturbing another variable/parameter r by a small amount, Δr, causes x^* to change by Δx^*. The relative (or normalized) change of r and x^* to this perturbation is $\frac{\Delta r}{r}$ and $\frac{\Delta x^*}{x^*}$, respectively. We can think of each of these as relative errors.

Definition: (sensitivity)

The sensitivity of an optimization problem ~~soln~~ (or model) solution x^* to another variable r is

$$S(x^*, r) = \frac{\Delta x^*/x^*}{\Delta r/r},$$

where Δx^* and Δr represent a relative change in x^* and r from their initial values, respectively. In the limit as $\Delta r \to 0$,

$$S(x^*, r) = \frac{dx^*}{dr} \cdot \frac{r}{x^*}$$

Back to the PC example: what is the sensitivity of D^* to r?

First, we need D^* as a function of r:

$$P(D, r) = 10{,}000(1 + rD)(250 - 100D) - 50{,}000$$

$$\frac{dP}{dD} = 10{,}000\left[(1 + rD)(-100) + r(250 - 100D)\right] = 0$$

$$\Rightarrow -200rD^* + 250r - 100 = 0$$

$$\Rightarrow D^* = \frac{5}{4} - \frac{1}{2r}$$

Then

$$S(D^*, r) = \frac{dD^*}{dr} \cdot \frac{r}{D^*} = +\frac{1}{2}r^{-2}\left(\frac{r}{D^*}\right) = \frac{1}{2rD^*}$$

(Figures)

So, for $r=0.5$ and $D^*=0.25$,

$$S(D^*, r) = \frac{1}{2(\frac{1}{2})(0.25)} = 4$$

We can also examine the sensitivity of P to r:

$$\left.\frac{dP}{dr}\right|_{D=0.25} = 10{,}000D(250-100D)\Big|_{D=0.25} = 2500(250-25)$$

$$= 2500(225)$$

$$S(P, r) = \frac{r}{P}(2500)(225) \approx 0.11$$

Robustness: A model with small sensitivities, or where variables/parameters with high sensitivities have a very low relative error, are said to be robust

Alternate derivation of sensitity: think of sensitivity as an error in x or y propagating through f:

$$\Delta f := f(x+\Delta x, y+\Delta y) - f(x,y)$$

$$\approx \frac{\partial f}{\partial x}(x,y)\Delta x + \frac{\partial f}{\partial y}(x,y)\Delta y \quad \text{by Taylor expansion}$$

Make relative error: $\frac{\Delta f}{f} = \left(\frac{\partial f}{\partial x} \cdot \frac{x}{f}\right)\frac{\Delta x}{x} + \left(\frac{\partial f}{\partial y} \cdot \frac{y}{f}\right)\frac{\Delta y}{y} \to S = \frac{\Delta f/f}{\Delta x/x}$

Chapter 2: Multivariable Optimization

Example 1

Case 2: Include the effect of advertizing

If we include the effect of advertizing, the model becomes

$$P(D,B) = \left[n(1+rD) + cB\right]\left[w - m - dD\right] - a - bB$$

The optimization problem is also now constrained:

$$\max_{B,D} \; P(D,B)$$

subject to $a + bB \in [0, 100{,}000] \Rightarrow B \in [0,10]$

Section 2.1: Unconstrained Optimization

As in single-variable calculus, extrema occur at critical points.

Definition:

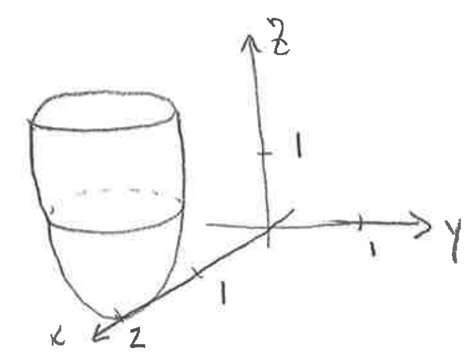
A function $f: \mathbb{R}^n \to \mathbb{R}$ has a critical point at $\vec{c}$ if $\nabla f = \vec{0}$ or undefined.

Note: The gradient of f is given by $\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right)$.

Example: $f(x,y) = (x-2)^2 + y^2$

$\nabla f = \left(2(x-2), 2y\right) = 0$

$\Rightarrow x = 2,\; y = 0 \quad \Rightarrow \vec{c} = (2,0)$



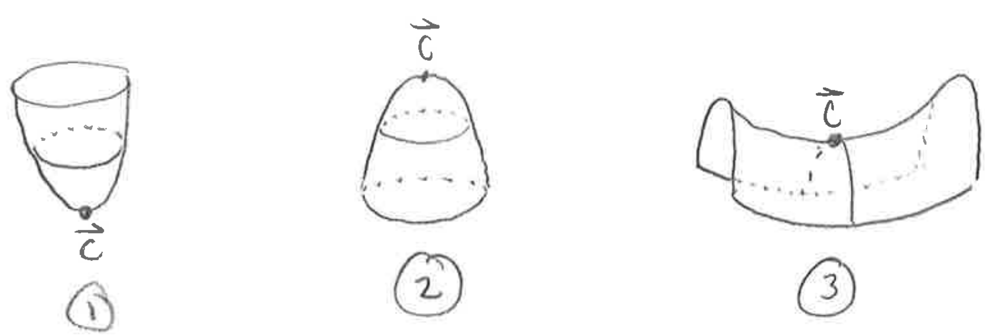
To determine if higher-dimensional critical points are maxima, minima, or neither, we need an equivalent to the second derivative test in 1D.

Reminder: the Hessian matrix is the matrix of all the 2^{nd} mixed partial derivatives of a function, $f: \mathbb{R}^n \to \mathbb{R}$. That is,

$$Hf = \begin{bmatrix} f_{x_1x_1} & f_{x_1x_2} & \cdots & f_{x_1x_n} \\ f_{x_2x_1} & f_{x_2x_2} & & \\ & & \ddots & \vdots \\ & & & f_{x_nx_n} \end{bmatrix}$$

Theorem:
Let $U \subseteq \mathbb{R}^n$ be open and $f: U \to \mathbb{R}$ be a $\mathcal{C}^2$ function. Suppose $\vec{c} \in U$ is a critical point of f

1. If $Hf(\vec{c})$ is positive definite, then f has a local min at $\vec{c}$
2. If $Hf(\vec{c})$ is negative definite, then f has a local max at $\vec{c}$
3. If $\det(Hf(\vec{c})) \neq 0$ but $Hf(\vec{c})$ is neither positive nor negative definite, then f has a saddle point at $\vec{a}$



So we can relate whether a critical point is a max/min to whether the associated Hessian matrix is positive or negative definite. We can now apply a result from linear algebra to make this result more practical. First, note that the sequence of principal minors of a matrix A refers to the sequence of the determinants of the upper leftmost square submatrices of A.

Second derivative test for local extrema

Given a critical point $\vec{c}$ of a function f that is C^2, consider $Hf(\vec{c})$ and the sequence of principle minors of $Hf(\vec{c})$, denoted by $d_1, d_2, \dots, d_n$, where $d_k = \det H_k f(\vec{c})$, and H_k is the upper leftmost $k \times k$ submatrix of $Hf(\vec{c})$.

1. If $d_k > 0$ for $k = 1, 2, \dots, n$, then f has a local min at $\vec{c}$
2. If $d_k < 0$ for k odd and $d_k > 0$ for k even, then f has a local max at $\vec{c}$

   (e.g., $-, +, -, +, \dots \Rightarrow$ neg. def)
3. If neither case 1 nor case 2 holds, then f has a saddle point at $\vec{c}$.

Ex: $f(x,y) = (x-2)^2 + y^2$

$\vec{c} = (2,0)$

$Hf(\vec{c}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \rightarrow d_1 > 0,\ d_2 > 0 \rightarrow$ local min at $\vec{c}$

Example: PC profit with discount and advertizing

Suppose for a moment that there is no constraint on advertizing. Let's find the maxima/minima of the profit function

$$P(D,B) = \left[10{,}000(1+rD) + 200B\right]\left[250 - 100D\right] - 50{,}000 - 10{,}000B$$

$$\frac{\partial P}{\partial D} = \underbrace{10{,}000}_{n} r(250 - 100D) - 100\left[\underbrace{10{,}000}_{n}(1+rD) + \underbrace{200}_{c}B\right]$$

$$\frac{\partial P}{\partial B} = \underbrace{200}_{c}(250 - 100D) - \underbrace{10{,}000}_{b} = 40{,}000 - 20{,}000D$$

Find critical points by solving $\frac{\partial P}{\partial D} = 0$ and $\frac{\partial P}{\partial B} = 0$. This is linear, so.

$$\begin{bmatrix} -100nr - 100nr & -100c \\ -100c & 0 \end{bmatrix}\begin{pmatrix} D \\ B \end{pmatrix} + \begin{pmatrix} 250nr - 100n \\ 250c - n \end{pmatrix} = 0$$

$$100cD = 250c - n \quad \rightarrow \quad D = \frac{250(200) - 10{,}000}{100(200)} = 2$$

$$200nrD + 100cB = 250nr - 100n$$

$$\rightarrow B = \frac{250nr - 100n - 400nr}{100c}$$

$$= \underbrace{\frac{n}{100c}}_{\frac{1}{2}}(-150r - 100) = -75r - 50$$

Next, we evaluate the Hessian at the critical point $(D^*, B^*) = (2, -75r - 50)$:

$$HP(D^*, B^*) = \begin{bmatrix} -200nr & -100c \\ -100c & 0 \end{bmatrix}$$ ← (no dependence on D^* or B^*)

3/5

Since the sequence of principal minors is

$d_1 < 0$, $d_2 < 0$,

$P(D^*, B^*)$ is a saddle point.

4/1

(Matlab interlude)

Example 2: TV sets

A manufacturer of color TV sets is planning the introduction of two new products, a 19–inch LCD flat panel set with a manufacturer's suggested retail price (MSRP) of $339 and a 21–inch LCD flat panel set with an MSRP of $399. The cost to the company is $195 per 19–inch set and $225 per 21–inch set, plus an additional $400,000 in fixed costs. In the competitive market in which these sets will be sold, the number of sales per year will affect the average selling price. It is estimated that for each type of set, the average selling price drops by one cent for each additional unit sold. Furthermore, sales of the 19–inch set will affect sales of the 21–inch set, and vice–versa. It is estimated that the average selling price for the 19–inch set will be reduced by an additional 0.3 cents for each 21–inch set sold, and the price for the 21–inch set will decrease by 0.4 cents for each 19–inch set sold. How many units of each type of set should be manufactured?

Variables:

s = number of 19–inch sets sold (per year)
t = number of 21–inch sets sold (per year)
p = selling price for a 19–inch set ($)
q = selling price for a 21–inch set ($)
C = cost of manufacturing sets ($/year)
R = revenue from the sale of sets ($/year)
P = profit from the sale of sets ($/year)

Step 3:

Assumptions:

$p = 339 - 0.01s - 0.003t$

$q = 399 - 0.004s - 0.01t$

$P = ps + qt - (400,000 + 195s + 225t)$

Objective: Maximize P

Step 4:

Find all critical points of P:

$$P(s,t) = (339 - 0.01s - 0.003t)s + (399 - 0.004s - 0.01t)t$$
$$- (400{,}000 + 195s + 225t)$$

$$\nabla P = \begin{pmatrix} \frac{\partial P}{\partial s} \\ \frac{\partial P}{\partial t} \end{pmatrix} = \begin{pmatrix} 339 - 195 - 0.02s - 0.007t \\ 399 - 225 - 0.007s - 0.02t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\rightarrow s = \frac{554{,}000}{117}, \quad t = \frac{824{,}000}{117}$$
$$\approx 4{,}735 \qquad \approx 7{,}043$$

$$\rightarrow P \approx 553{,}641$$

How do we handle constraints?

1. Compact Region Method ← similar to Closed Interval Method in 1D
2. Substitution
3. Lagrange multipliers

① Compact Region Method

Heine Borel Theorem:

A subset $S \subseteq \mathbb{R}^n$ is compact iff it's closed and bounded.

Note: Closed: includes the boundary
Bounded: doesn't go to ∞

Compact Region Method:

Given a function $f: \mathbb{R}^n \to \mathbb{R}$ on a compact subsets of $\mathbb{R}^n$:

① Optimize on the inside of S:

$\nabla f = \vec{0}$, 2^{nd} deriv test

② optimize on the boundary

Parameterize and plug into the fuction

③ The largest value from ①,② <u>must</u> be the absolute maximum on S.

(similar for minimum)

5/1

<u>EX</u>: Find absolute extrema of $f(x) = x^2 - y^2$ on $D = \{x^2 + y^2 \leq 1\}$

$\nabla f = (2x, -2y) \rightarrow$ crit pt at $(0,0)$

$Hf = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \rightarrow \begin{matrix} d_1 > 0 \\ d_2 < 0 \end{matrix} \rightarrow$ saddle point

Boundary: $(\cos t, \sin t)$, $t \in [0, 2\pi]$

$f|_{\partial D} = f(\cos t, \sin t) = \cos^2 t - \sin^2 t$

max: $t = 0, \pi \leftrightarrow (x,y) = (1,0), (-1,0) \rightarrow f = 1$

min: $t = \frac{\pi}{2}, \frac{3\pi}{2} \leftrightarrow (x,y) = (0,1), (0,-1) \rightarrow f = -1$

$\therefore$ Abs max is $f=1$ at $(1,0), (-1,0)$

Abs min is $f=-1$ at $(0,1), (0,-1)$

Section 2.2: Lagrange Multipliers

We'll start our discussion of Lagrange multipliers by considering problems with equality constraints:

$$\underset{\vec{x} \in \mathbb{R}^n}{\text{maximize}} \quad f(\vec{x})$$

$$\text{subject to} \quad g_1(\vec{x}) = 0$$
$$g_2(\vec{x}) = 0$$
$$\vdots$$
$$g_N(\vec{x}) = \vec{0}$$

<u>Example</u>: Find abs max/min of $f(x,y) = x^2 + y^2$ on $y^2 = x+5$

Lagrange multipliers: find x, y such that

$$\nabla f = \lambda \nabla g$$
$$g(x,y) = c$$

for some λ (the Lagrange multiplier) and constant c.

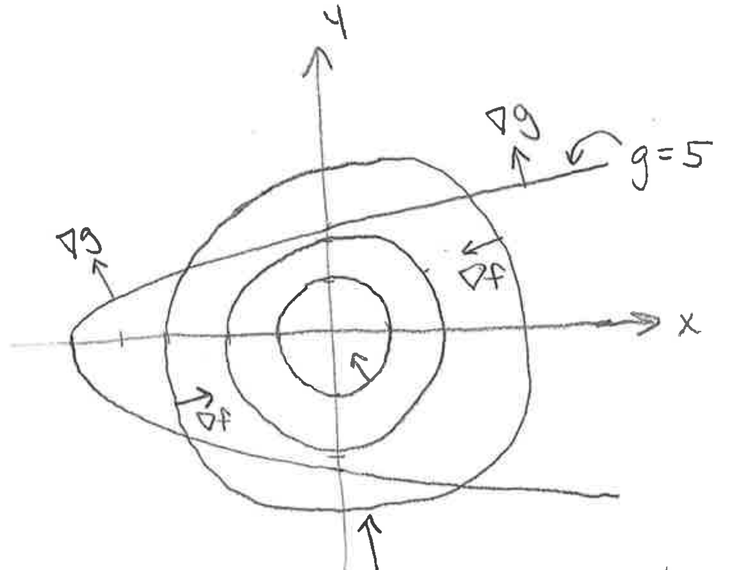
Rephrase problem: let $g(x,y) = y^2 - x = 5$

$$\begin{cases} \nabla(x^2+y^2) = \lambda \nabla(y^2 - x) \\ y^2 - x = 5 \end{cases}$$

What are we doing geometrically?

(Mathematica figure)

5/3

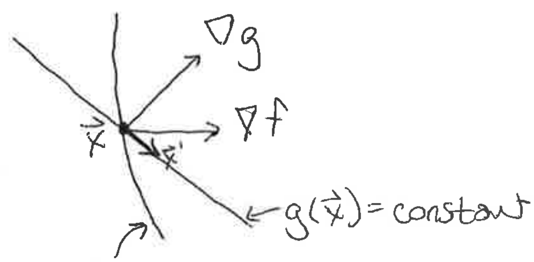


level curves/isoclines of f: $f(x,y) = \text{constant}$

want to find where $\nabla f \parallel \nabla g$

The gradient of a function points in the direction of the greatest rate of increase of that function.

If ∇f and ∇g are not parallel at some point $\vec{x}$:



then moving in the direction of $\vec{x}'$ changes $f(x,y)$ while maintaining $g(\vec{x}) = \text{constant}$.

Back to example: solve

$$\begin{cases} \nabla(x^2+y^2) = \lambda \nabla(y^2 - x) \\ y^2 - x = 5 \end{cases}$$

$$\Rightarrow \begin{pmatrix} 2x \\ 2y \end{pmatrix} = \lambda \begin{pmatrix} -1 \\ 2y \end{pmatrix} \longrightarrow \begin{matrix} 2x = -\lambda \\ 2y = 2\lambda y \quad (*) \end{matrix}$$

$$y^2 - x = 5$$

Either $\lambda=1$ or $y=0$, by $(*)$

$$\lambda = 1 \Rightarrow x = -\tfrac{1}{2}$$

$$y^2 + \tfrac{1}{2} = 5 \Rightarrow y = \pm\frac{3}{\sqrt{2}} \qquad \Rightarrow \left(-\tfrac{1}{2}, \pm\frac{3}{\sqrt{2}}\right)$$

$$y = 0 \Rightarrow x = -5 \quad \Rightarrow (-5, 0)$$

In general:

Let $\vec{x}^*$ be a solution of $\min\limits_{\vec{x}\in\mathbb{R}^n} f(\vec{x})$ (or max) subject

to

$$g_1(\vec{x}) = c_1$$
$$g_2(\vec{x}) = c_2$$
$$\vdots$$
$$g_N(\vec{x}) = c_N$$

Then, if all functions are differentiable and the vectors $\nabla g_1(\vec{x}^*), \nabla g_2(\vec{x}^*), \ldots, \nabla g_N(\vec{x}^*)$ are linearly independent, there exist Lagrange multipliers $\lambda_1, \lambda_2, \ldots, \lambda_N$ s.t.

$$\nabla f(\vec{x}^*) = \lambda_1 \nabla g_1(\vec{x}^*) + \cdots + \lambda_N \nabla g_N(\vec{x}^*)$$

__Example:__ PC profit

$$\max_{B,D} P(D,B)$$

subject to $B \in [0,10] \longleftarrow$ constraint: $B \le 10$, $B \ge 0$

Note: $\nabla P \neq 0$ on $B \in [0,10]$

Let's try one constraint at a time:

$$g_1(D,B) = 10 = B$$

$$\rightarrow \nabla P = \lambda(0,1)$$

$$\rightarrow \frac{\partial P}{\partial D} = 0 = nr\,(250-100D) - 100\,(n(1+rD)+cB)$$

$$\frac{\partial P}{\partial B} = \lambda = 200(250-100D) - 10{,}000$$

$$\rightarrow D = \frac{250}{100} - \frac{\lambda + 10{,}000}{200(100)}$$

$$\rightarrow 250 - 100D = \frac{\lambda + 10{,}000}{200}$$

Plug into $\frac{\partial P}{\partial D}$:

$$nr\left(\frac{\lambda+10{,}000}{200}\right) - 100\,(n(1+rD)+c\underbrace{B}_{=10}) = 0$$

$$\rightarrow \lambda = 39{,}000 > 0$$

$$D = 0.05$$

Next, try $g_2(D,B) = B = 0$ $\rightarrow \lambda = \frac{\partial P}{\partial B} > 0$

$$\rightarrow D = 0.25, \quad \lambda > 0$$

Since $\lambda > 0$, the max must be at the top constraint. If the max was at $B=0$, then $\frac{\partial P}{\partial B} < 0$ at $B=0$, so we would get $\lambda < 0$.

$\therefore$ $D = 0.05$ and $B = 10$ give maximum, $P = 2{,}851{,}250$

(Matlab figure)

Sensitivity: Is D as significant as B?

Recall:

$$S(P,r) = \frac{\partial P}{\partial r}\frac{r}{P} \quad , \quad S(D,r) = \frac{\partial D}{\partial r}\frac{r}{D}$$

we need the optimal D as a function of the constraint, B_{max}:

$$\frac{\partial P}{\partial D} = 0 \longrightarrow 40{,}000 - 20{,}000 D = \lambda B_{max}$$

$$2 - D = \frac{\lambda B_{max}}{20{,}000}$$

$$D = 2 - \frac{\lambda B_{max}}{20{,}000}$$

Then

$$\frac{\partial P}{\partial B_{max}} = \frac{\partial P}{\partial D}\frac{\partial D}{\partial B_{max}} + \frac{\partial P}{\partial B}\frac{\partial B}{\partial B_{max}}$$

$$= \nabla P \cdot \left(\frac{\partial D}{\partial B_{max}}, \frac{\partial B}{\partial B_{max}}\right)$$

$$= \lambda \nabla g \cdot \left(\frac{\partial D}{\partial B_{max}}, \frac{\partial B}{\partial B_{max}}\right)$$

$$= \lambda (0,1) \cdot \left(\frac{\partial D}{\partial B_{max}}, \frac{\partial B}{\partial B_{max}}\right)$$

$$= \lambda$$

$$= 39{,}000$$ ← the L.M of the maximum

$$\Rightarrow S(P, B_{max}) = \lambda \frac{B_{max}}{P} \approx 0.137$$

Real-world interpretation of λ: 'Shadow price' - the change in P associated with the change in the constraint. In this case, a 1% increase in B corresponds (to 1st order) to a ~0.137% increase in profit

<u>Problem:</u> How to deal with inequality constraints?

<u>Example:</u> TV sets

$$\left.\begin{aligned} p &= 300 - 0.03s - 0.005t \\ q &= 400 - 0.005s - 0.03t \\ C &= 400{,}000 + 150s + 200t \end{aligned}\right\} \underline{P} = ps + qt - C$$

$$\max_{(s,t)} \quad P(s,t)$$

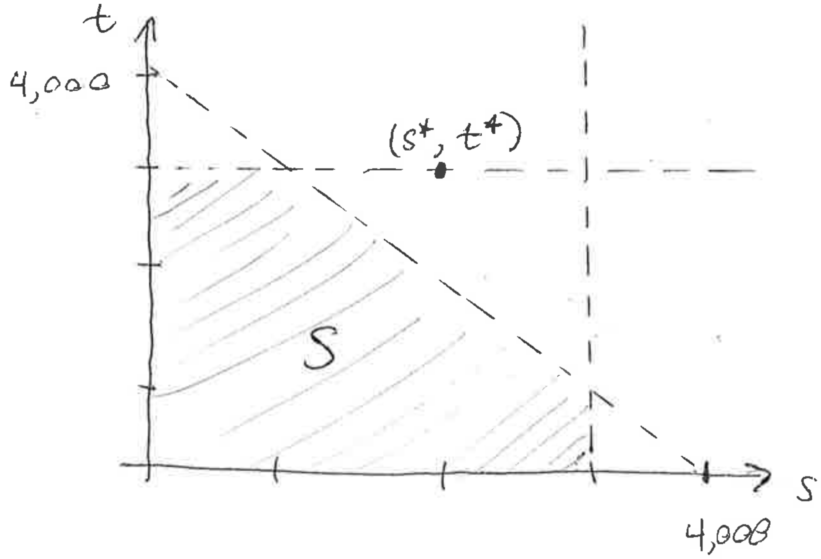
subject to
$$s \geq 0$$
$$t \geq 0$$
$$s \leq 3{,}000$$
$$t \leq 3{,}000$$
$$s + t \leq 4{,}000$$

<u>Step 1:</u> Ignore all inequality constraints and determine all critical points of P:

$$\nabla P = \begin{pmatrix} \frac{\partial P}{\partial s} \\ \frac{\partial P}{\partial t} \end{pmatrix} = \begin{pmatrix} 300 - 0.06s - 0.01t - 150 \\ 400 - 0.01s - 0.06t - 200 \end{pmatrix} = 0$$

$$\rightarrow s^* = 2{,}000, \quad t^* = 3{,}000, \quad P^* = 50{,}000$$

Step 2: Consider the feasible set/region, S



Note: $\nabla P \neq 0$ inside S. $\therefore$, the optimum must lie on a boundary of S!

Step 3: Add an appropriate equality constraint to the (unconstrained) optimization problem

Consider the condition $s+t \leq 4000$:

solve $\nabla P = \lambda \nabla g$

$g(s,t) = s+t = 4000$

$$\Rightarrow \begin{pmatrix} 300 - 0.06s - 0.01t - 150 \\ 400 - 0.01s - 0.06t - 200 \\ s+t-4000 \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} 0.06s + 0.01t + \lambda \\ 0.01s + 0.06t + \lambda \\ s + t \end{pmatrix} = \begin{pmatrix} 150 \\ 200 \\ 4000 \end{pmatrix}$$

$\Rightarrow \quad s = 1500, \quad t = 2{,}500, \quad \lambda = 35$

$\Rightarrow \quad P^* = 32{,}500$

Check all inequality constraints: they are satisfied.

To show that this is the absolute maximum, we would need to optimize P over all line segments.

Section 2.3: Sensitivity Analysis

Sensitivity analysis:
What is the sensitivity of the optimal profit P^* to the constraint $s+t = C = 4{,}000$?

Theorem:
Let $\vec{x}^* \in \mathbb{R}^n$ be the solution of

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

$$\text{subject to } g(\vec{x}) = c$$

Let f^* be the corresponding optimal value and λ the corresponding Lagrange multiplier. Then,

$$S(f^*, c) = \lambda \frac{c}{f^*}$$

Pf:

$$\frac{df^*}{dc} \overset{\text{chain rule}}{=} \nabla f(\vec{x}^*) \cdot \frac{d\vec{x}^*}{dc} = \lambda \nabla g(\vec{x}^*) \cdot \frac{d\vec{x}^*}{dc} \overset{\text{chain rule}}{=} \lambda \frac{dg}{dc} = \lambda$$

$$\Rightarrow \quad S(f^*, c) = \frac{\partial f^*}{\partial c} \frac{c}{f^*} = \lambda \frac{c}{f^*}$$

$\square$

Back to example: $S(P^*, c) = 35\left(\frac{4{,}000}{32{,}500}\right) \approx 4.3$

Definition:

The Lagrange multiplier λ is also called a shadow price (or marginal cost) of the constraint.

Note: $\frac{\Delta P}{\Delta C} = \lambda \Rightarrow \Delta P = \lambda \Delta C$

$\Rightarrow$ For a change ΔC we expect an increase in profits of $\lambda \Delta C$.

Example: Multiple Lagrange multipliers

$$\min_{(x,y,z)} \quad f(x,y,z) = (x+1)^2 + y^2 + z^2$$

subject to $g_1(x,y,z) = x^2+y^2+z^2 = 4$

$g_2(x,y,z) = x+y+z = 2$

Formulate as a Lagrange multiplier problem bey introducing λ_1, λ_2 for g_1, g_2:

$$\begin{cases} \nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 \\ g_1(x,y,z) = 0 \\ g_2(x,y,z) = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \begin{pmatrix} 2(x+1) \\ 2y \\ 2z \end{pmatrix} = \lambda_1 \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ x^2+y^2+z^2 = 4 \\ x+y+z = 2 \end{cases}$$

$$\Rightarrow \begin{cases} 2x+z = 2\lambda_1 x + \lambda_2 & (1) \\ 2y = 2\lambda_1 y + \lambda_2 & (2) \\ 2z = 2\lambda_1 z + \lambda_2 & (3) \\ x^2+y^2+z^2=4 & (4) \\ x+y+z=2 & (5) \end{cases}$$

Express x, y, z in terms of λ_1, λ_2:

$(1) \Rightarrow \quad x = \dfrac{\lambda_2 - 2}{2-2\lambda_1}$

$(2) \Rightarrow \quad y = \dfrac{\lambda_2}{2(1-\lambda_1)}$

$(3) \Rightarrow \quad z = \dfrac{\lambda_2}{2(1-\lambda_1)}$

Then

$(5) \Rightarrow (\lambda_2 - 2) + \lambda_2 + \lambda_2 = 2(2)(1-\lambda_1)$

$\Rightarrow \lambda_2 = \dfrac{4(1-\lambda_1)+2}{3}$

and

$(4) \Rightarrow$ a quadratic equation in λ_1

$\Rightarrow \lambda_1 = \frac{1}{2}, \frac{3}{2}$

Case 1: $\lambda_1 = \frac{1}{2}, \lambda_2 = \frac{4}{3}$

$\Rightarrow \vec{x}^* = \left(-\frac{2}{3}, \frac{4}{3}, \frac{4}{3}\right) \quad \Rightarrow f(\vec{x}^*) = \frac{11}{3}$

Thus, $\left(-\frac{2}{3}, \frac{4}{3}, \frac{4}{3}\right)$ is the global minimum

Case 2: $\lambda_1 = \frac{3}{2}, \lambda_2 = 0$

$\Rightarrow \vec{x}^* = (2, 0, 0) \quad \Rightarrow f(\vec{x}^*) = 9$

This process gets tedious and challenging very quickly as we add more constraints. Is there a way to find $f(x^*)$ numerically?

Here's another motivating example:

Example: Newspaper problem from 1.4.9.
Maximize profit with respect to the subscription price.
The profit function is

$$N(R) = \underbrace{s}_{\#\text{ subscribers}}\ \overbrace{q}^{\text{profit/subscriber}}$$

$$s = (n_0 - nR) \quad , \quad q = p_0 + rR$$

where

$n_0 = 80{,}000 = \#$ current subscribers

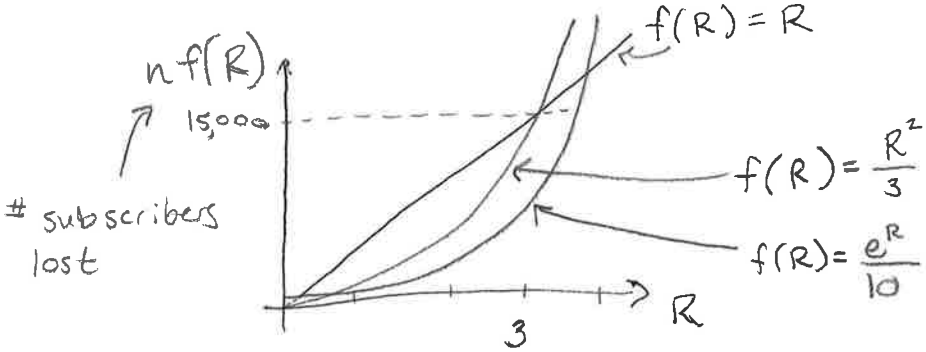
$n = 5{,}000 = \#$ subscribers lost if price raised \$0.10

$r =$ estimated raise of price $= \$0.10$

$R =$ amt of estimation to raise the price

$p_0 = \$1.5 =$ current subscription price

We'd like to investigate the robustness of our optimum price to the form of the model. What if the assumption about n was based on a survey? Perhaps we might expect a different relationship between R and n



$f(R) = \frac{R^2}{3}$: For a price increase < \$.30, not much effect on subscriptions (relative to a linear scaling), but a greater effect above \$0.30

$f(R) = \frac{e^R}{10}$: Small price increase causes some subscriptions lost, but then little effect until reach a 'critical value' around $3.6 = R$

We could conceive of multiple options, but how does our optimum price change with each? $P^* := p_0 + rR^*$

Case 1: $f(R) = R$

From HW1 solutions, $S(P^*, n) \approx -0.52$

Case 2: $f(R) = \frac{R^2}{3}$

$$S(P, n) = \frac{dP}{dn}\left(\frac{n}{P}\right)$$

$$N(R) = \left(n_0 - n\frac{R^2}{3}\right)(p_0 + rR)$$

$$\frac{dN}{dR} = r\left(n_0 - n\frac{R^2}{3}\right) - \frac{2nR}{3}(p_0 + rR) = 0$$

Case 3: $f(R) = \frac{e^R}{10}$

$$\frac{dN}{dR} = r\left(n_0 - n\frac{e^R}{10}\right) - \frac{n}{10}e^R\left(p_0 + rR\right) = 0$$

We need to find this zero numerically.

Chapter 3: Computational Methods for Optimization

3.1: One variable optimization

In one-variable opt, optimizing is equivalent to a root-finding problem. I.e.,

$$\max_x f(x^*) \longleftrightarrow f'(x^*) = 0$$

(finding the critical point(s))

There are a few ways we can solve this root-finding problem numerically.

Newton's Method

If $F(x)$ is a (well-behaved) differentiable function, we can approximate it by a line, close to our estimate of the critical point, x_i (so $F(x) = f'(x)$ from above)

$$F(x^*) = F(x_i) + F'(x_i)(x^* - x_i)$$

we want $F'(x^*) = 0$, so we set

$$F(x_i) + F'(x_i)(x^* - x_i) = 0$$

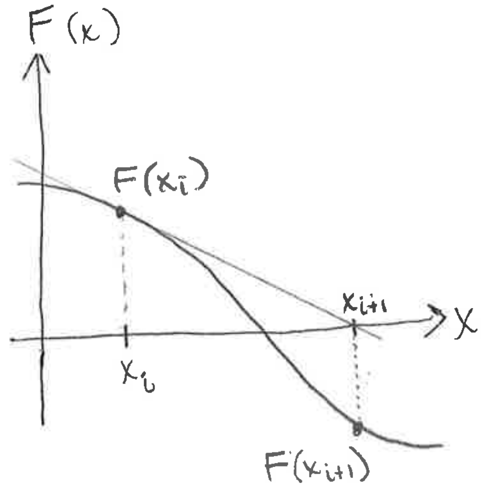
and solve to find x^*, the location of the root:

$$F(x_i) = F'(x_i)(x_i - x^*) \rightarrow \frac{F(x_i)}{F'(x_i)} = x_i - x^*$$

$$\rightarrow x^* = x_i - \frac{F(x_i)}{F'(x_i)}$$

This x^* becomes our new estimate, x_{i+1}, for the critical point of $f(x)$, i.e. the point where $f'(x) = F(x) = 0$

The idea:



In terms of the original function $f(x)$, the update step is

$$x_{i+1} = x_i - \frac{f'(x)}{f''(x)},$$

where x_{i+1} is the updated (improved) guess for x^*, where $f(x^*)=0$

Bisection method

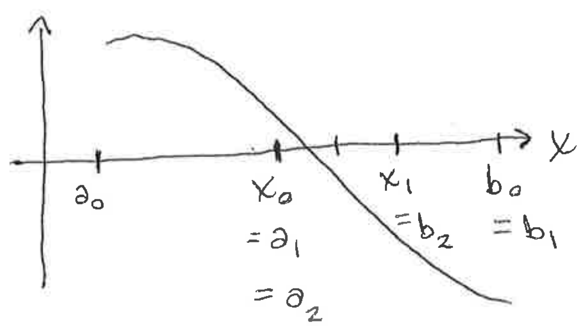
Again, consider $F(x) = f'(x)$.

Method: Start with an interval $[a_0, b_0]$ with $F(a_0) > 0$ and $F(b_0) < 0$ (or $F(a_0) < 0$ and $F(b_0) > 0$).

Let $x_0 = \frac{a_0 + b_0}{2}$,

If $F(x_0) > 0$, define new interval $[x_0, b_0] = [a_1, b_1]$

Otherwise, use $[a_1, b_1] = [a_0, x_0]$.



$F(x_0) > 0$

$F(x_1) < 0$

8/3

<u>Regula falsi</u> (aka false position method)

We can improve the bisection method by using

$$x_i = \frac{a_i F(b_i) - b_i F(a_i)}{F(b_i) - F(a_i)}$$

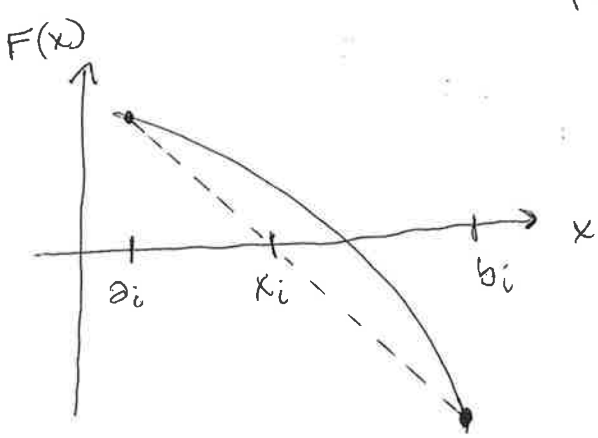
instead of $x_i = \frac{a_i + b_i}{2}$. This method <u>always converges</u> (unlike Newton's Method), is faster than Bisection Method, and sometimes is faster than other methods too, so it's a good method to keep in mind (e.g. for when Newton's doesn't converge or converges very slowly).

Idea: $$\underbrace{F(x_0)}_{=0} - F(b_0) = \frac{F(b_0) - F(a_0)}{b_0 - a_0}(x_0 - b_0)$$

$$x_0 = b_0 - F(b_0)\frac{b_0 - a_0}{F(b_0) - F(a_0)}$$

$$= \frac{\cancel{b_0 F(b_0)} - b_0 F(a_0) - \cancel{F(b_0) b_0} + a_0 F(b_0)}{F(b_0) - F(a_0)}$$

$$= \frac{a_0 F(b_0) - b_0 F(a_0)}{F(b_0) - F(a_0)}$$



Then proceed as in Bisection Method:

$F(x_i) > 0 \Rightarrow [x_i = a_{i+1}, b_i = b_{i+1}]$

$F(x_i) < 0 \Rightarrow [a_i = a_{i+1}, x_i = b_{i+1}]$

8/4

(Matlab interlude)

Pseudocode for Newton's Method:

Let $x(n)$ = approximate location of root after n iterations

N = number of iterations

Input: $x(0)$, N

Process: for $n = 1$ to N

$$x(n) \leftarrow x(n-1) - F(x(n-1)) / F'(x(n-1))$$

end

Output: $x(N)$

Try these three methods to find the ~~critical~~ roots ~~points~~ of $f(x) = x^3 + 2x + 1$

3.2: Multivariable Optimization

Random search

This method randomly selects N feasible points within a feasible region you define. The smallest (or largest) of these is assigned as the smallest (or largest) value of the objective function.

Starting value: $\vec{x}_0$

Iteration: $(i \to i+1)$

Randomly select N new candidates $\vec{w}_K$ in a "neighborhood" of $\vec{x}_i$ and set $\vec{x}_{i+1}$ to the optimal $\vec{w}_K$:

$$\vec{x}_{i+1} = \underset{\vec{x} \in \{\vec{w}_1, \dots, \vec{w}_N\}}{\operatorname{argmin}} f(\vec{x})$$

Newton's Method

Interpretation: minimize $f(\vec{x}) \Rightarrow$ find $\vec{x}^*$ s.t. $\nabla f(\vec{x}^*) = 0$

Idea: let $\vec{F}(\vec{x}) = \nabla f(\vec{x})$

Taylor expand near an initial guess, $\vec{x}_0$:

$$\vec{F}(\vec{x}_0 + \vec{x}') = \vec{F}(\vec{x}_0) + D\vec{F}(\vec{x}_0)\vec{x}' + \mathcal{O}(\|\vec{x}'\|^2)$$

where

$$D\vec{F}(\vec{x}_0) = \left.\begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial F_n}{\partial x_1} & \cdots & \frac{\partial F_n}{\partial x_n} \end{pmatrix}\right|_{\vec{x} = \vec{x}_0}$$

Choose $\vec{x}'$ such that $\vec{F}(\vec{x}_0) + D\vec{F}(\vec{x}_0)\vec{x}' = 0$:

$$\iff D\vec{F}(\vec{x}_0)\vec{x}' = -\vec{F}(\vec{x})$$

$$\iff \vec{x}' = -\left(D\vec{F}(\vec{x}_0)\right)^{-1}\vec{F}(\vec{x})$$

So, with a starting value of $\vec{x}_0$, we can estimate where $\vec{F}(\vec{x}) = 0$ by iterating using

$$\vec{x}_{i+1} = \vec{x}_i - \left(D\vec{F}(\vec{x})\right)^{-1}\vec{F}(\vec{x})$$

In Matlab: $D\vec{F}(\vec{x})\backslash\vec{F}(\vec{x})$ better than $\texttt{inv}\left(D\vec{F}(\vec{x})\right)*\vec{F}(\vec{x})$

↑ mldivide

$$= \vec{x}_i - \left(Hf(\vec{x})\right)^{-1}\nabla f(\vec{x})$$

2D example:

Solving $\nabla f(x,y) = 0$, we start with an initial guess, (x_0, y_0)

$$0 = f_x(x^*, y^*) \equiv F(x^*, y^*) = F(x_0, y_0) + F_x(x_0, y_0)(x^* - x_0) + F_y(x_0, y_0)(y^* - y_0)$$

$$0 = f_y(x^*, y^*) \equiv G(x^*, y^*) = G(x_0, y_0) + G_x(x_0, y_0)(x^* - x_0) + G_y(x_0, y_0)(y^* - y_0)$$

$$\Rightarrow \begin{bmatrix} F_x & F_y \\ G_x & G_y \end{bmatrix} \begin{pmatrix} x^* - x_0 \\ y^* - y_0 \end{pmatrix} = \begin{pmatrix} -F(x_0, y_0) \\ -G(x_0, y_0) \end{pmatrix}$$

$$\begin{pmatrix} x^* - x_0 \\ y^* - y_0 \end{pmatrix} = \frac{1}{F_x G_y - G_x F_y} \begin{bmatrix} G_y & -F_y \\ -G_x & F_x \end{bmatrix} \begin{pmatrix} -F(x_0, y_0) \\ -G(x_0, y_0) \end{pmatrix}$$

$$= \frac{-1}{F_x G_y - G_x F_y} \begin{pmatrix} G_y F - F_y G \\ G_x F - F_x G \end{pmatrix}$$

$$\Rightarrow \quad x_{i+1} = x_i - \frac{f_{yy}f_x - f_{xy}f_y}{f_{xx}f_{yy} - f_{xy}f_{yx}}$$

$$y_{i+1} = y_i - \frac{f_{xx}f_y - f_{yx}f_x}{f_{xx}f_{yy} - f_{yx}f_{xy}}$$

This is equivalent to the n-D case we just derived using the derivative matrix.

Example: Minimize $f(x,y) = 2x^2 - (x-1)(y-2)^2 + 2y^4$

$$\nabla f = \begin{pmatrix} 4x - (y-2)^2 \\ -2(x-1)(y-2) + 8y^3 \end{pmatrix}$$

$$Hf = \begin{bmatrix} 4 & -2(y-2) \\ -2(y-2) & -2(x-1) + 24y^2 \end{bmatrix}$$

Iteration: $\begin{cases} \vec{x}_0 = (x_0, y_0) \\ \vec{x}_{i+1} = \vec{x}_i - (Hf(\vec{x}_i))^{-1} \nabla f(\vec{x}_i) \end{cases}$

⚠ Beware of false solutions: this example has 3 critical points:

$(1,0)$

$(0.52, 0.55)$ ← local min

$(2.18, -0.95)$ ← global min

Newton's Method for constrained optimization problems

Example: (from day 7)

minimize $f(x,y,z) = (x+1)^2 + y^2 + z^2$

subject to $g_1 = x^2+y^2+z^2 = 4$

$g_2 = x+y+z = 2$

(i)

Solve the system

$$2x + 2 - 2\lambda_1 x - \lambda_2 = 0$$

$$2y - 2\lambda_1 y - \lambda_2 = 0$$

$$2z - 2\lambda_1 z - \lambda_2 = 0$$

$$x^2+y^2+z^2-4 = 0$$

$$x+y+z-2 = 0$$

write this as $\vec{F}(\vec{x}) = \vec{0}$, $\vec{x} = (x, y, z, \lambda_1, \lambda_2)$

$$\Rightarrow D\vec{F}(\vec{x}) = \begin{bmatrix} 2-2\lambda_1 & 0 & 0 & -2x & -1 \\ 0 & 2-2\lambda_1 & 0 & -2y & -1 \\ 0 & 0 & 2-2\lambda_1 & -2z & -1 \\ 2x & 2y & 2z & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

(really Hf)

Iterate: $\begin{cases} \vec{x}_0 \\ \vec{x}_{i+1} = \vec{x}_i - (D\vec{F}(\vec{x}))^{-1}\vec{F}(\vec{x}) \end{cases}$

Section 3.3: Linear Programming

For when both the objective and constraint functions are linear.

Def:
A linear optimization/programming problem in standard form is as follows:

Given $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, $x \in \mathbb{R}^n$

maximize $\quad f(\vec{x}) = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$

$$\text{subject to } \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \leq b_1 \\ \vdots \qquad\qquad\qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \leq b_n \end{cases}$$

and $x_1 \geq 0$, $x_2 \geq 0$, ..., $x_n \geq 0$

Example: (3.4 in text)

A farm has 625 acres available for planting corn, wheat, & oats
1,000 acre-ft of water available for irrigation,
300hr of labor per week

Requirements for corn, wheat, oats:

per acre:	corn	wheat	oats
irrigation	3.0	1.0	1.5
labor	0.8	0.2	0.3
yield ($)	400	200	250

Variables: x_1 = acres of corn

x_2 = acres of wheat

x_3 = acres of oats

y = total yield (\$)

Assumptions:

$$\begin{cases} 3.0x_1 + 1.0x_2 + 1.5x_3 \leq 1000 & \text{(irrigation)} \\ 0.8x_1 + 0.2x_2 + 0.3x_3 \leq 300 & \text{(labor)} \\ x_1 + x_2 + x_3 \leq 625 & \text{(acres)} \\ x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0 \\ y = 400x_1 + 200x_2 + 250x_3 \end{cases}$$

Objective: Maximize y

For linear programs, the optimal point will always occur at one of the corner points of the feasible region.

Theorem:

Suppose the feasible set S of a linear optimization problem is nonempty (and convex). Then, there is a solution $\vec{x}^*$ to the optimization problem that is a corner point of S.

More precisely, exactly n of the $m+n$ constraints

$$a_{i1}x_1 + \cdots + a_{in}x_n \leq b_i \quad , \quad i = 1, \ldots, m$$

$$x_j \geq 0 \quad , \quad j = 1, \ldots, n$$

are binding, i.e. fulfilled with '='.

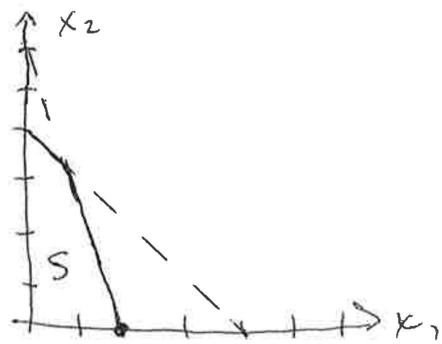
For simple cases, we can use this to directly guess the correct solution:

Example: $\max\limits_{x_1, x_2 \in \mathbb{R}} \quad f(x_1, x_2) = 2x_1 + x_2$

subject to $\quad x_1 + x_2 \leq 4$

$3x_1 + x_2 \leq 6$

Feasible set:



corner $(2,0)$ with binding constraints

$3x_1 + x_2 = 6$

$x_2 = 0$

$(f(x_1, x_2) = 4)$

+ corner $(1,3)$ with binding constraints

$$\left.\begin{aligned} 3x_1 + x_2 &= 6 \\ x_1 + x_2 &= 4 \end{aligned}\right\} \to f(x_1, x_2) = 5$$

⋮

After checking all corners: global maximum at $(1,3)$ with $f(x_1, x_2) = 5$.

For more complex problems, we use the simplex method.

Idea of the simplex method:

Motivation: number of corners $\sim \binom{n+m}{n}$

$\rightarrow$ for $n=50, m=100$: $\binom{150}{50} = \frac{150!}{50!\,100!}$

$\approx 2 \cdot 10^{60}$

we

Idea: introduce slack variables $x_{n+1}, \dots, x_{n+m}$ and write side conditions as equality constraints:

$$\begin{cases} a_{11}x_1 + \dots + a_{1n}x_n + x_{n+1} = b_1 \\ a_{21}x_1 + \dots + a_{2n}x_n \qquad + x_{n+2} = b_2 \\ \vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n \qquad\qquad + x_{n+m} = b_m \end{cases}$$

where $x_1 \geq 0, \dots, x_{n+m} \geq 0$.

(Matlab first)

Algorithm:

Initial value: let $x_1 = x_2 = \dots = x_n = 0$

$x_{n+1} = b_1, \dots, x_{n+m} = b_m$

Iterate: In every step, go from one corner $(x_1^k, \dots, x_{n+m}^k)$ to another $(x_1^{k+1}, \dots, x_{n+m}^{k+1})$ by selecting an "inactive" index j (with $x_j^k = 0$) and an "active" index* i ($x_i^k \geq 0$ allowed) and exchanging them

Motivation: computing $\vec{x}^{k+1}$ from $\vec{x}^k$ is cheap!

Example: $\left\{\begin{array}{ll} x_i^0 = 0 & \text{for } i = 1, \dots, n \\ x_{n+j}^0 = b_j & \text{for } j = 1, \dots, m \end{array}\right\}$ is a corner

Exchange $i = n+2$ and $j = 1$:

$$\left\{\begin{array}{l} a_{11} x_1^1 + x_{n+1}^1 = b_1 \\ a_{21} x_1^1 \qquad\quad = b_2 \\ \qquad * \end{array}\right.$$

← (everything else stays the same)

This means: $x_{n+2}^1 = 0$

$$x_1^1 = \frac{b_2}{a_{21}}$$

$$x_2^1 = \dots \quad (*)$$

In Matlab, this can be implemented using linprog (linprog uses the simplex method for the 'dual' problem by default, but we'll use it anyway)

Example: (planting problem)

max $\quad y = 400x_1 + 200x_2 + 250x_3$

s.t.

$$3.0x_1 + 1.0x_2 + 1.5x_3 \leq 1000 \quad ①$$

$$0.8x_1 + 0.2x_2 + 0.3x_3 \leq 300 \quad ②$$

$$x_1 + x_2 + x_3 \leq 625 \quad ③$$

$$x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0$$

In Matlab: min f

s.t. $Ax \leq b$

So $-y = -(400x_1 + 200x_2 + 250x_3)$

$\rightarrow f = (-400, -200, -250)$

$$A = \begin{bmatrix} 3.0 & 1.0 & 1.5 \\ 0.8 & 0.2 & 0.3 \\ 1 & 1 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

$$b = \begin{bmatrix} 1000 \\ 300 \\ 625 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Result: $y = 162{,}500$ with $x_1 = 187.5$

$x_2 = 437.5$

$x_3 = 0$

constraints 1 and 3 are binding

Back to example: $\max \ y = 400x_1 + 200x_2 + 250x_3$

$$\begin{cases} 3.0x_1 + 1.0x_2 + 1.5x_3 \leq 1000 = c & (1) \\ 0.8x_1 + 0.2x_2 + 0.3x_3 \leq 30 & (2) \\ x_1 + x_2 + x_3 \leq 625 & (3) \\ x_1 \geq 0, \ x_2 \geq 0, \ x_3 \geq 0 \end{cases}$$

binding

we got $y = 162500$, $x_1 = 187.5$, $x_2 = 437.5$, $x_3 = 0$.

How would we do a sensitivity analysis?

$$S(y,c) = \frac{dy}{dc}\frac{c}{y}$$

Fortunately, we already know how to do this - Lagrange formalism! Consider our example with the binding constraints:

$$\nabla y = \lambda_1 \nabla g_1 + \lambda_3 \nabla g_3$$

$$\Rightarrow \begin{pmatrix} 400 \\ 200 \\ 250 \end{pmatrix} = \lambda_1 \begin{pmatrix} 3.0 \\ 1.0 \\ 1.5 \end{pmatrix} + \lambda_3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\Rightarrow \lambda_1 = 100, \ \lambda_3 = 100$$

First constraint: $\frac{dy}{dc} = \lambda_1 = 100$

$\Rightarrow$ an increase of water supply by 1 acre-ft results in an increase in revenue of \$100

Algorithm example:

$$\min(\overbrace{-x_1 \mp 2x_2}^{f})$$

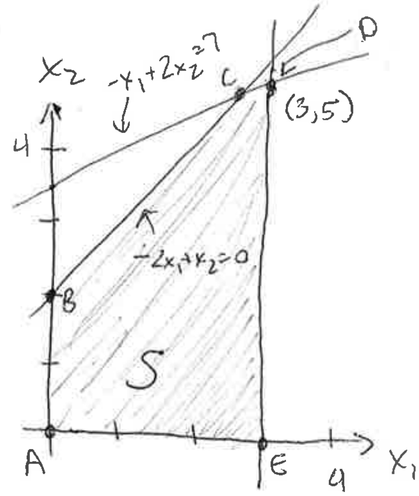
s.t. $-2x_1 + x_2 + s_3 = 2$

$-x_1 + 2x_2 + s_4 = 7$

$x_1 + s_5 = 3$

$x_1, x_2 \geq 0$

$s_3, s_4, s_5 \geq 0$



($s_i \rightarrow$ slack variables; replace $s_i \rightarrow x_i$)

We know we will find the maximum at a corner point (also sometimes called an extreme point), so we choose one feasible solution (that satisfies the constraints) that is a corner point, and see if we can do better.

Let's try a simple case: take the slack variables to equal the constraints, so that $x_1, x_2 = 0$. (starting at A)

Note: we need at least 3 of the x_i's to not equal 0 (we call those the basic variables) to satisfy the constraints:

$\Rightarrow x_3 = 2,\ x_4 = 7,\ x_5 = 3$ (basic variables)

$\Rightarrow x_1 = x_2 = 0$ (non-basic variables)

$= \min(f) = 0$

Now let's see if we can do better by switching one of the non-basic variables with one of the basic variables.

Notice that f is improved if we increase either x_1 or x_2. f will be improved more if we choose to change x_2.

Note: we can't change both x_1 + x_2 because we want to stay on a constraint.

$$x_2 \neq 0 \longrightarrow \{x_2, x_4, x_5\} \text{ can satisfy the constraints.}$$
$$\{x_2, x_3, x_5\}$$
$$\{x_2, x_3, x_4\}$$
$$\vdots$$

Let's look at $\{x_2, x_3, x_4\}$.
Moving along the constraint $x_1 = 0$ (increasing x_2), which constraint do we hit first by decreasing x_3 or x_4?

$$-2(0) + x_2 + x_3 = 2$$
$$-(0) + 2x_2 + x_4 = 7$$
$$x_1 + x_5 = 3$$

Decreasing x_3 works better.
$\Rightarrow$ $\underline{x_2 = 2}$, $\underline{x_3 = 0}$, and $x_4 \geq 0$ (point B)
we want to maximize x_2 (to minimize f)

We have changed our basic variables:

$$x_2, x_4, x_5 \neq 0$$
$$x_1, x_3 = 0$$

$$-2(0) + x_2 + 0 = 2 \rightarrow x_2 = 2$$
$$-0 + 2x_2 + x_4 = 7 \rightarrow x_4 = 3$$
$$(0) + x_5 = 3 \rightarrow x_5 = 3$$

This step is called a <u>pivot</u>

We repeat the procedure until we can no longer improve f:

1. Check to see possible improvements to f by changing one non-basic variable
2. Choose direction to switch
3. Pivot (switch basic, non-basic variables) and repeat

Summary:

First step: $(x_1 = 0, x_2 = 0)$ $\min(f) = 0$

Basic vars	Non-basic vars
x_3, x_4, x_5	x_1, x_2

The problem is

$$\min(-x_1 - 2x_2)$$

$$\text{s.t. } -2x_1 + x_2 + x_3 = 2$$

$$-x_1 + 2x_2 + x_4 = 7$$

$$x_1 + x_5 = 3$$

Second step: $(x_1 = 0, x_2 = 2)$ $\min(f) = -4$

Basic vars	Non-basic vars
x_2, x_4, x_5	x_1, x_3

We change the objective function in terms of non-basic variables

$$\min(-x_1 \underbrace{-4x_1 + 2x_3 - 4}_{\text{from } -2x_1 + x_2 + x_3 = 2 \;\rightarrow\; x_2 = 2 + 2x_1 - x_3,\; -2x_2 = -4 - 4x_1 + 2x_3}) = \min(-5x_1 + 2x_3 - 4)$$

$$\text{s.t. } -2x_1 + x_2 + x_3 = 2$$

$$-x_1 + 2x_2 + x_4 = 7$$

$$x_1 + x_5 = 3$$

The coeff of x_1 in obj. fun is greater, so we change x_1

$\rightarrow$ affects x_2, x_4, x_5

At $x_1 = 3$, $x_5 = 0$, which is off-constraint

Moving along the first constraint (recall $x_3 = 0$),

we have $x_2 = 2 + 2x_1$

$\cancel{x_1 = 1 - \frac{1}{2}x_2}$ $\rightarrow -x_1 + 4x_1 + 4 + x_4 = 7$ ~~[illegible]~~

$\rightarrow 3x_1 = 3 - x_4$

$\rightarrow \underline{x_1 = 1}, \underline{x_4 = 0}, x_2 = 4$

x_1 maximized if $x_4 = 0$

<u>third step</u>: $(x_1 = 1, x_2 = 4)$ $\min(f) = -5$

<u>basic</u>	<u>non-basic</u>
x_1, x_2, x_5	x_3, x_4

New obj fun:

$$f = -9 + \frac{5}{3}x_4 - \frac{4}{3}x_3$$

since $\begin{cases} x_2 = 2 + 2x_1 - x_3 & \text{from (1)} \\ -x_1 + 2(2 + 2x_1 - x_3) + x_1 = 7 & \text{from above \& (3)} \end{cases}$

$$\rightarrow \begin{cases} x_1 = 1 + \frac{2}{3}x_3 - \frac{1}{3}x_4 \\ x_2 = 2 + 2\left(1 + \frac{2}{3}x_3 - \frac{1}{3}x_4\right) - x_3 \end{cases}$$

s.t. $-2x_1 + x_2 + x_3 = 2$

$-x_1 + 2x_2 + x_4 = 7$

$x_1 + x_5 = 3$

Fourth step: $(x_1 = 3,\ x_2 = 5)$ $\min(f) = -13$

basic	non-basic
x_1, x_2, x_3	x_4, x_5

$$f = 2x_5 + x_4 - 13$$

The obj. fun can't be improved by increasing the ~~B~~ non-basic variables.

$\therefore$ The minimum is $f = -13$, with binding conditions giving the extreme point.

Section 3.4: Discrete Optimization

An integer programming problem is a linear programming problem in which the decision variables are constrained to take integer values

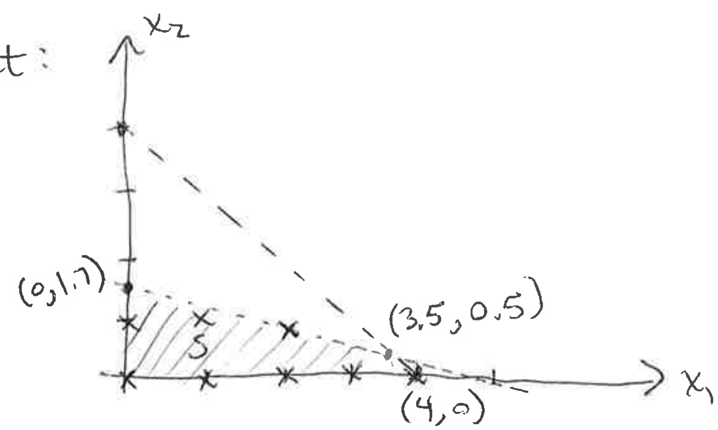
Example: $\max_{x_1, x_2 \in \mathbb{Z}} f = x_1 + 2x_2$

s.t. $x_1 + x_2 \leq 4$

$x_1 + 3x_2 \leq 5$

$x_1, x_2 \geq 0$

Feasible set:



Without the integer constraint, the max occurs at $(3.5, 0.5)$

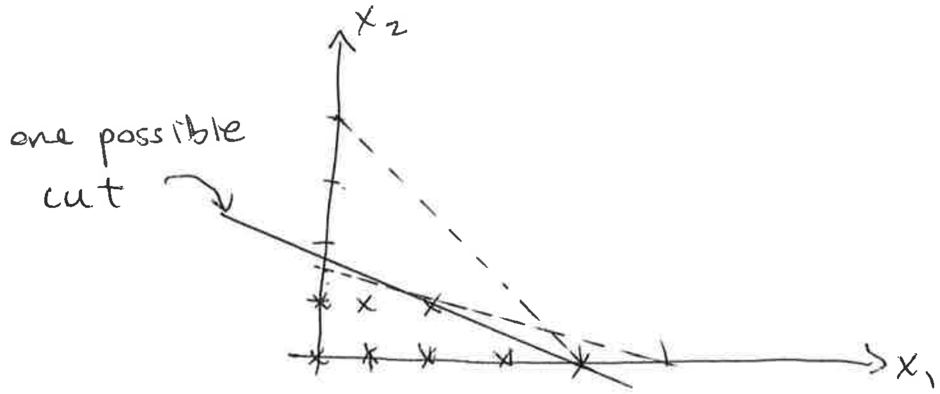
Strategies: ("bundle" methods)

Gomory's cut

If soln not an integer point, introduce an additional constraint ("cut") that removes this point from the feasible set.

* Need to still keep all integer points in the feasible set

12/3



<u>Branch-and ~~cut~~ bound:</u>

Exclude non-integer optimum by dividing up ("branching") the problem. E.g.,

$$x_1 \le 3 \quad \text{or} \quad x_1 \ge 4$$

leads to two new problems with these as additional constraints.

Solve both branched problems. If integer solution, use this as the new candidate. If non-integer solution, branch again.

<u>Other methods:</u>

- combinatorial methods
- dynamic programming

<u>Example:</u> (warehouse

y_i ~ 1 if opened, 0 if closed

x_{ij} ~ amount to be sent from warehouse i to ~~warehouse~~ customer j

f_i ~ fixed operating costs

c_{ij} ~ per unit transportation cost

d_j ~ demand per customer j

$d_1 = 100$
$d_2 = 300$
$d_3 = 1000$

$f_1 = 5{,}000$
$f_2 = 10{,}000$
$f_3 = 8{,}000$
$f_4 = 3{,}000$

c_{ij}	1	2	3
1	10	10	10
2	20	30	40
3	35	50	100
4	10	100	100

minimize $\sum_{i=1}^{4} \sum_{j=1}^{3} c_{ij} x_{ij} + \sum_{i=1}^{4} f_i y_i$

s.t. $y_i = 0$ or 1

$x_{ij} \geq 0$

$\sum_{i=1}^{4} x_{ij} = d_j$ ← meet demand

$\sum_{j=1}^{3} x_{ij} - y_i \left(\sum_{j=1}^{3} d_j\right) \leq 0$ ← only operate if $y_i = 1$