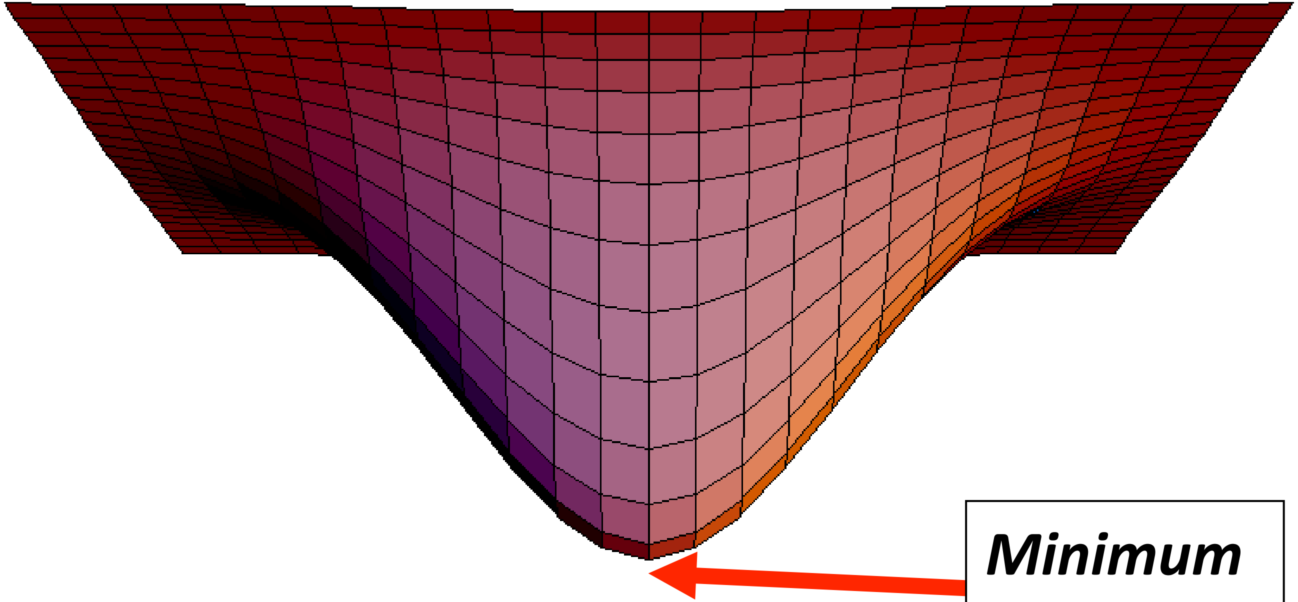
15.053x Week 6

Optimization Methods in Business Analytics

Nonlinear Programming

Nonlinear Optimization



Minimum value

Thanks to Rob Freund and Juan Carlos Ferrer for sharing their slides

Overview

- **Portfolio optimization.**
- **Convex functions and convex sets.**
- **Local optimization vs. global optimization.**
  - **Convex minimization problems are "easy."**
- **More applications of nonlinear programming**

Portfolio optimization 1: Intro

Portfolio optimization

- Important topic in Finance
- 1990 Nobel Prize in Economics
  - Miller, Sharpe, and Markowitz
- Knowledge about probability and random variables will be useful, but it is not required for these videos.

Portfolio optimization 1

Investment choice: 3 stocks.

Data: expected returns, standard deviations (5 yr)

Assumption (for now): stock returns are independent of each other (uncorrelated).

Company	Five Year Expected Return (%)	Standard Dev. of Return (%)
IBM	65	15
Prudential	80	18
Raytheon	50	10

Company	Five Year Expected Return (%)	Standard Dev. of Return (%)
IBM	65	15
Prudential	80	18
Raytheon	50	10

65% expected return: "on average," for every dollar invested, the investor will have $1.65 after five years.

15% standard deviation: can be roughly interpreted as follows. The "return" of $1 invested after five years is likely (approximately 2/3 chance) to be between 65% - .15% and 65% + .15%, (50% and 80%)

Objectives

- Create a portfolio
- Maximize expected return
- Minimize risk (standard dev. of portfolio)

Assumption.   Investments with a higher expected return and lower “risk” are preferable.

This assumption is not always true in practice.

Which lottery would you prefer?

Lottery 1.

- **A 90% chance of winning $1,000,000**
- **A 10% chance of losing $100,000**

Lottery 2.

- **A 50% chance of winning $1,000,000**
- **A 50% chance of winning $100,000**

Portfolio optimization 2: A model

The three random variables

X_I = 5-year return on investing in IBM

X_P = 5-year return on investing in Prudential

X_R = 5-year return on investing in Raytheon

X_I, X_P, X_R are random variables.

$E[X_I] = 65$	$SD[X_I] = 15$
$E[X_P] = 80$	$SD[X_P] = 18$
$E[X_R] = 50$	$SD[X_R] = 10$

Expected values

Suppose X and Y are any random variables.

Then $E(X + Y) = E(X) + E(Y)$.

Standard deviations and variances

Suppose that X is a random variable.

- Let VAR(X) denote Variance of X
- $\mathrm{VAR}(X) = \mathrm{SD}(X)^2$ $\mathrm{SD}(X) = \mathrm{VAR}(X)^{1/2}$
  - The math of variances is simpler than that of standard deviations

Suppose X and Y are independent random variables.

Then $\mathrm{VAR}(X + Y) = \mathrm{VAR}(X) + \mathrm{VAR}(Y)$.

- Not true for standard deviations. $\mathrm{SD}(X + Y) \neq \mathrm{SD}(X) + \mathrm{SD}(Y)$.
- Not true for dependent random variables.

The decision variables

I = proportion of portfolio invested in IBM.

P = proportion of portfolio invested in Prudential

R = proportion of portfolio invested in Raytheon

$I + P + R = 1$

$I \geq 0, P \geq 0, R \geq 0$

Returns and expected returns

W = return of the entire portfolio

$= I\, X_I + P\, X_P + R\, X_R$

$E[W] = I * E(X_I) + P * E(X_P) + R * E(X_R)$

$= 65\, I + 80\, P + 50\, R$

Standard deviation and variance

$SD(P\,X_P) = SD(X_P) \times P = 18\,P$

$VAR(P\,X_P) = VAR(X_P) \times P^2 = 18^2\,P^2$

$VAR(W)$ = Variance of the entire portfolio

$= VAR(X_I) \times I^2 + VAR(X_P) \times P^2 + VAR(X_R) \times R^2$

$VAR(W) = 225\,I^2 + 324\,P^2 + 100\,R^2$

$SD(W) = VAR(W)^{1/2}$.

The portfolio optimization model

$$\min \quad VAR(W) = 225\, I^2 + 324\, P^2 + 100\, R^2$$

$$\text{s.t.} \quad E(W) = 65\, I + 80\, P + 50\, R \geq 65$$

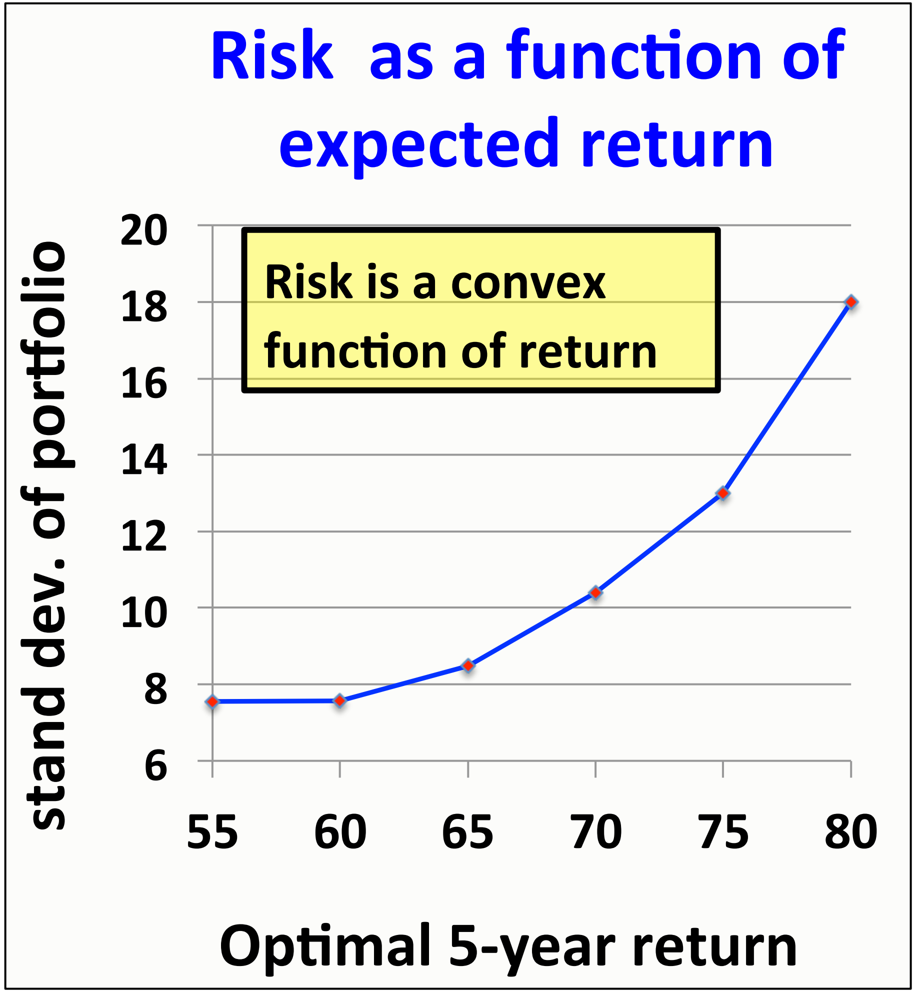
$$I + P + R = 1$$

$$I \geq 0, \quad P \geq 0, \quad R \geq 0$$

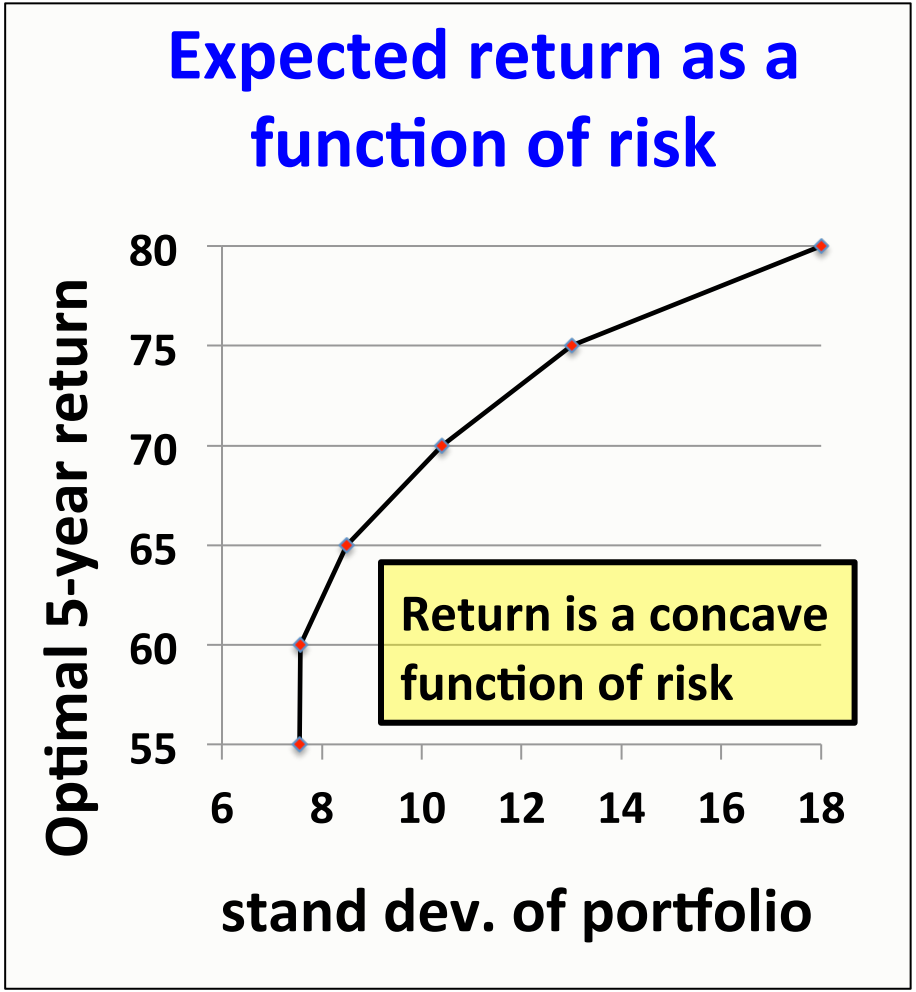
"65" is the target value.

We would optimize for multiple target values.

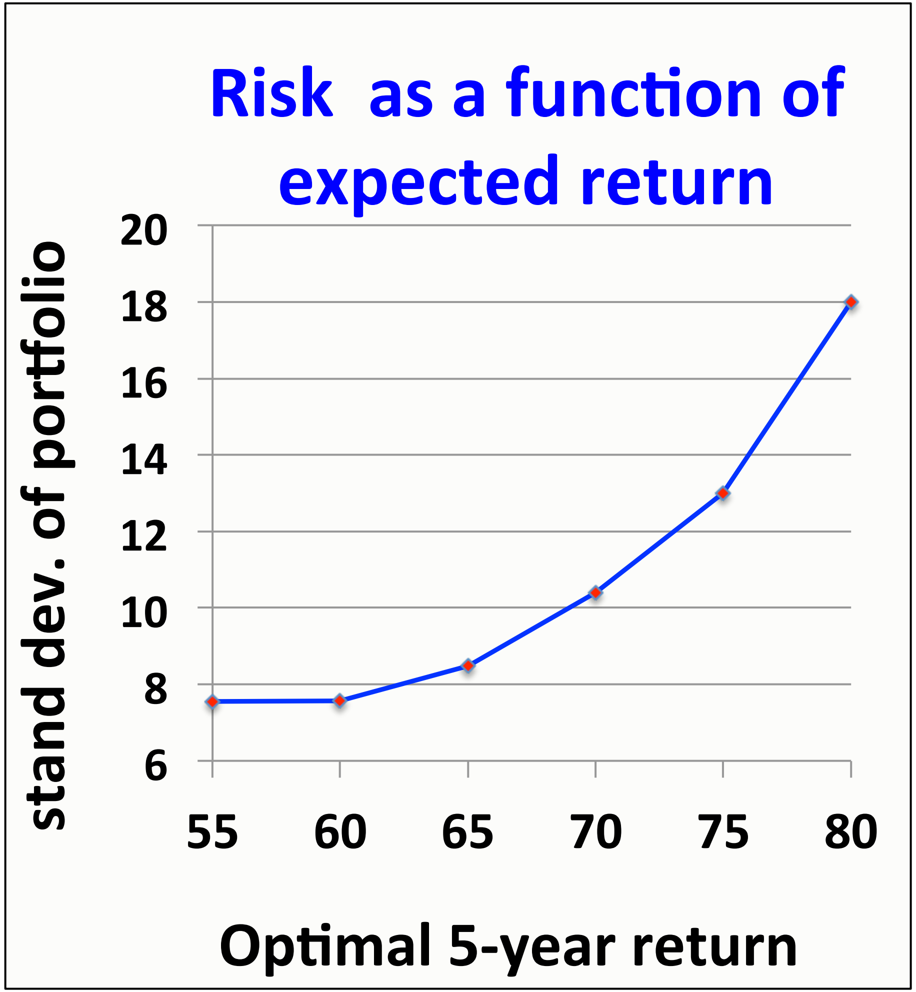
Portfolio optimization 3: Analysis

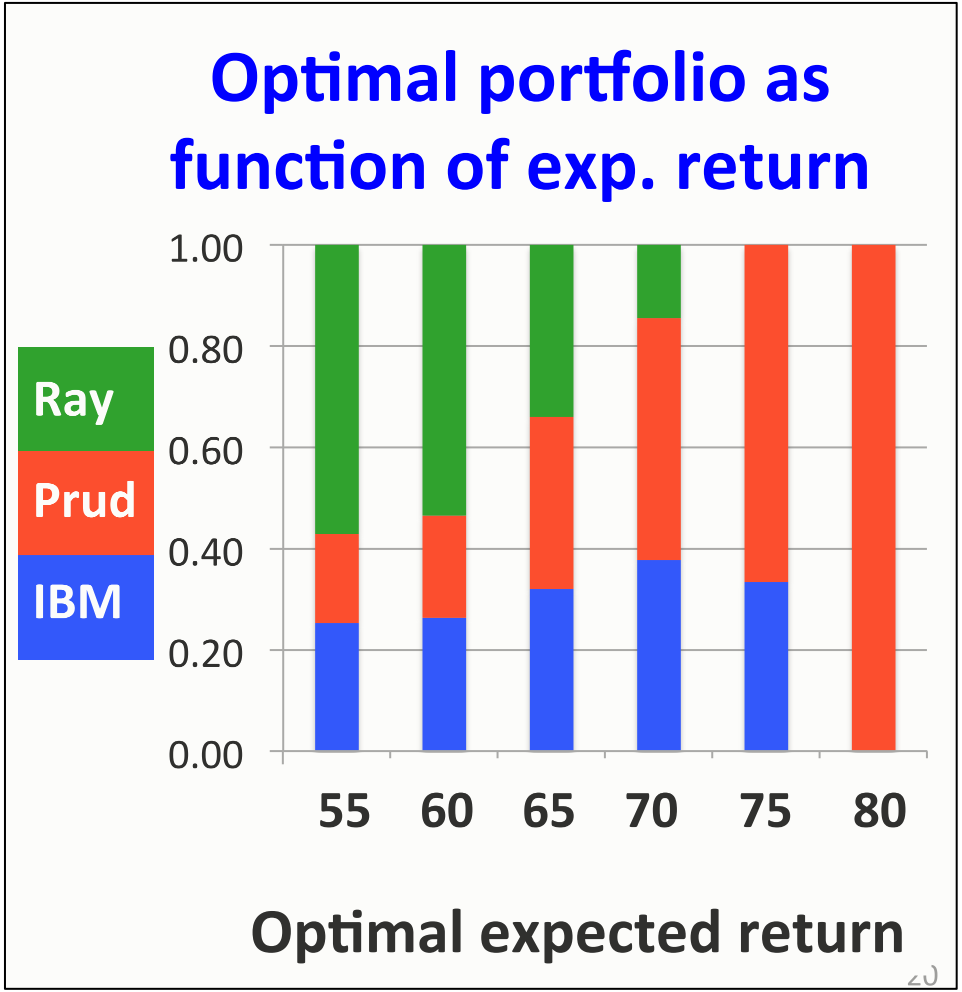


Risk as a function of expected return
Risk is a convex function of return
stand dev. of portfolio
20
18
16
14
12
10
8
6
55
60
65
70
75
80
Optimal 5-year return



Expected return as a function of risk
Return is a concave function of risk
Optimal 5-year return
80
75
70
65
60
55
6
8
10
12
14
16
18
stand dev. of portfolio


Risk as a function of expected return
stand dev. of portfolio
20
18
16
14
12
10
8
6
55
60
65
70
75
80
Optimal 5-year return


Optimal portfolio as function of exp. return
Ray
Prud
IBM
1.00
0.80
0.60
0.40
0.20
0.00
55
60
65
70
75
80
Optimal expected return

The SR Report

Name	Final Value	Reduced Gradient
IBM	0.320	0
Prudential	0.340	0
Raytheon	0.340	0

Name	Final Value	Lagrange Multiplier
100% invested	1.00	-185.80
return >= R	65.00	5.08

The reduced cost is called a <u>reduced gradient</u> in a nonlinear program.

The shadow price is called a <u>Lagrange multiplier</u> in a nonlinear program.

These are derivatives (slopes).

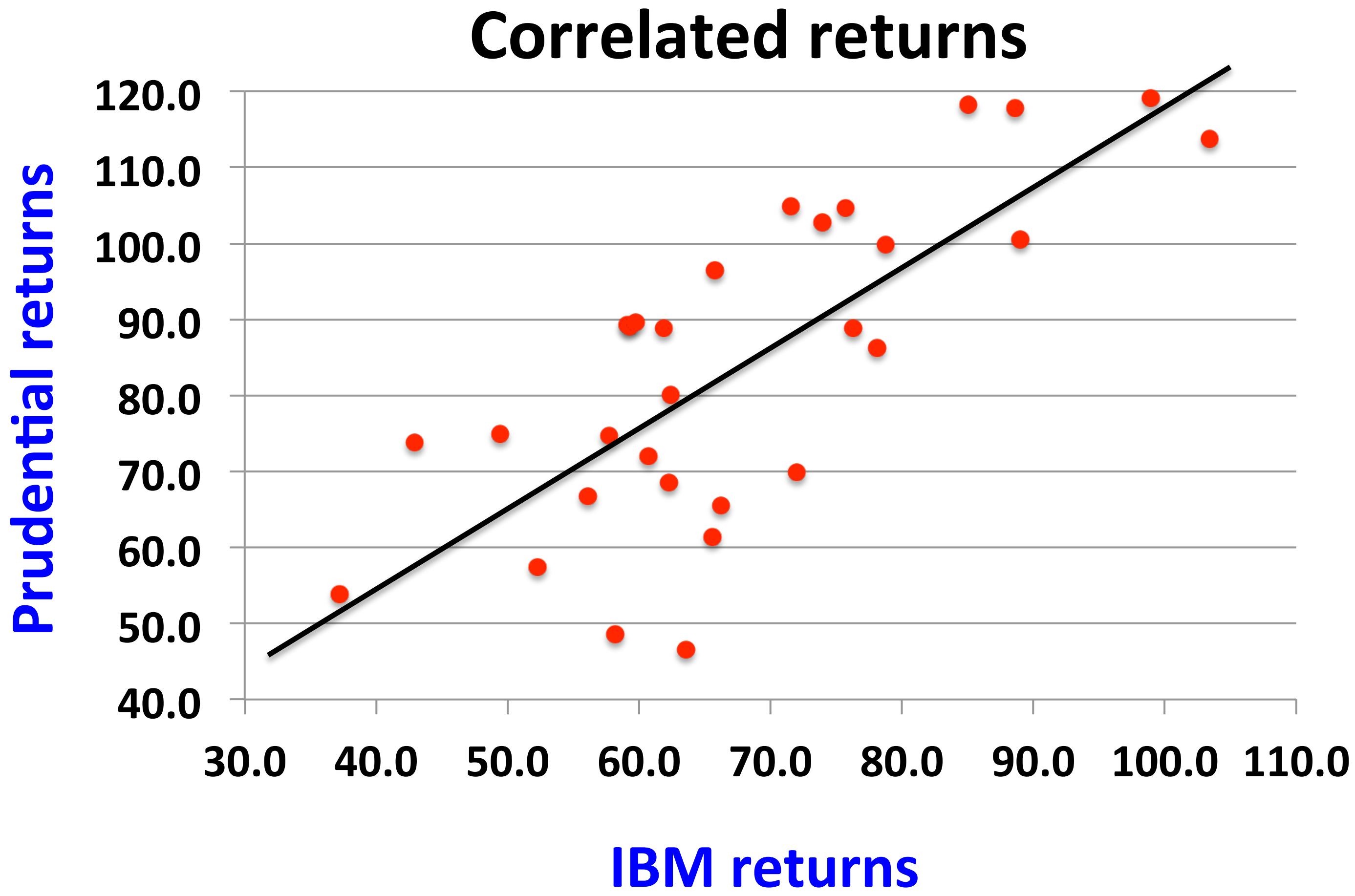
They are most accurate for VERY small changes in the RHS.

Portfolio optimization 4:
Correlated returns

30 simulations of 5 years

Uncorrelated returns

30 simulations of 5 years



What happens if RVs are not independent?

- The updated model is the same as before except that the formula for the VAR(W) is modified.

	IBM	Prudential	Raytheon
IBM	225	150	-75
Prudential	150	324	60
Raytheon	-75	60	100

Covariance Matrix

Var(W) if the RVs are not independent.

$$VAR(W) = \sum_{i \in \{I,P,R\}} \sum_{j \in \{I,P,R\}} COV(i,j) \times i \times j$$

$$= 225\ I^2 + 324\ P^2 + 100\ R^2 +$$

$$+ (2 \times 150 \times I \times P) + (2 \times -75 \times I \times R) + (2 \times 60 \times P \times R)$$

	IBM	Prudential	Raytheon
IBM	225	$\boxed{150}$	-75
Prudential	150	324	60
Raytheon	-75	60	100

Covariance Matrix

The portfolio optimization model

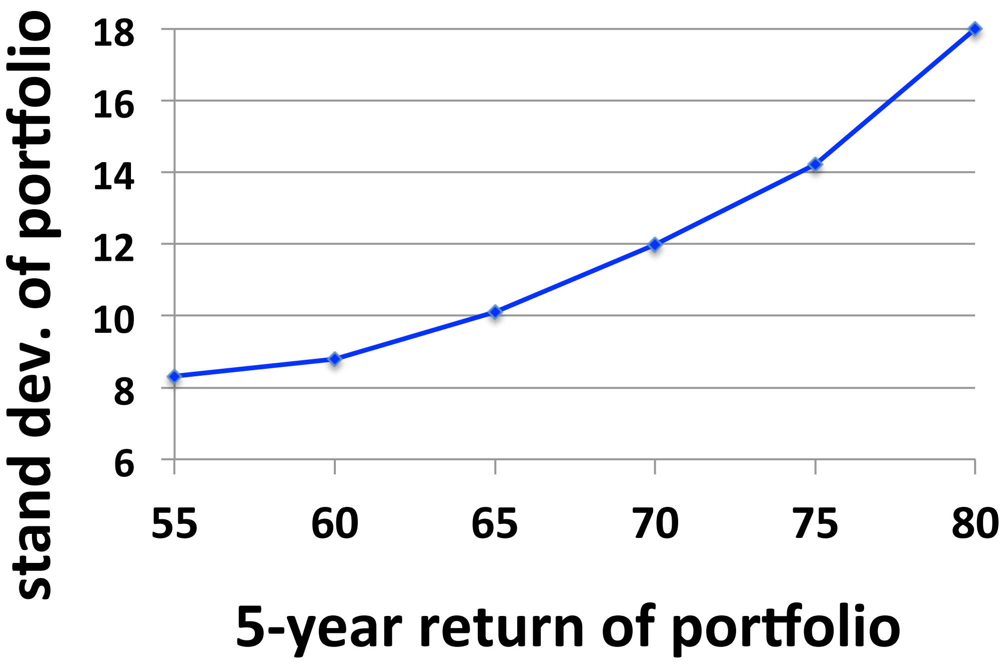
$$\min \quad VAR(W) = 225\, I^2 + 324\, P^2 + 100\, R^2 + $$
$$+ \; (2 \times 150 \times I \times P) + (2 \times -75 \times I \times R) + (2 \times 60 \times P \times R)$$

$$\text{s.t.} \quad E(W) = 65\, I + 80\, P + 50\, R \geq \text{Target}$$

$$I + P + R = 1$$

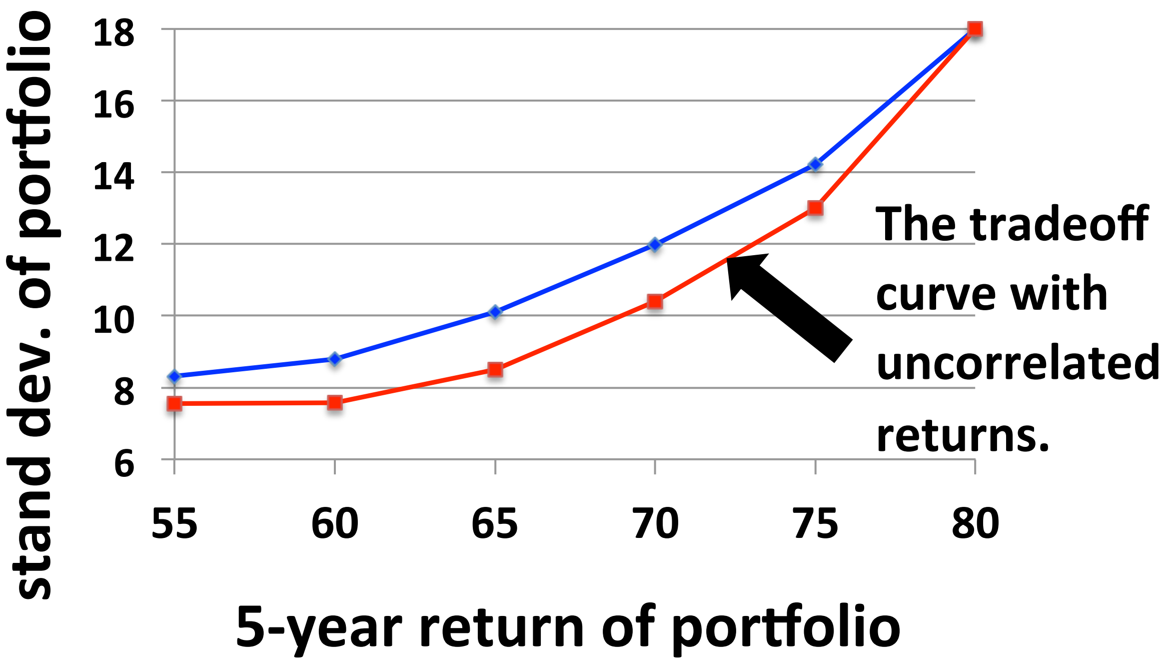
$$I \geq 0, \quad P \geq 0, \quad R \geq 0$$

Risk vs Return (correlated returns)



Risk vs Return

Independent vs correlated returns



To achieve the same return with positively correlated variables, the risk increases.

Portfolio optimization 5:

A common error in estimating "returns"

On estimating returns

We don't know the expected returns or the covariance matrix. <u>We estimate them from data</u>.

Company	Five Year Expected Return (%)	Standard Dev. of Return (%)
IBM	65	15
Prudential	80	18
Raytheon	50	10

	IBM	Prudential	Raytheon
IBM	225	150	-75
Prudential	150	324	60
Raytheon	-75	60	100

The estimate of expected return for each stock is unbiased.

It is equally likely to be too high or too low.

The expected return of the best performing stock is very biased.

Suppose we select the stock with the best return.

Our estimate of this stock is VERY biased.

The expected return in reality is much lower than our estimate.

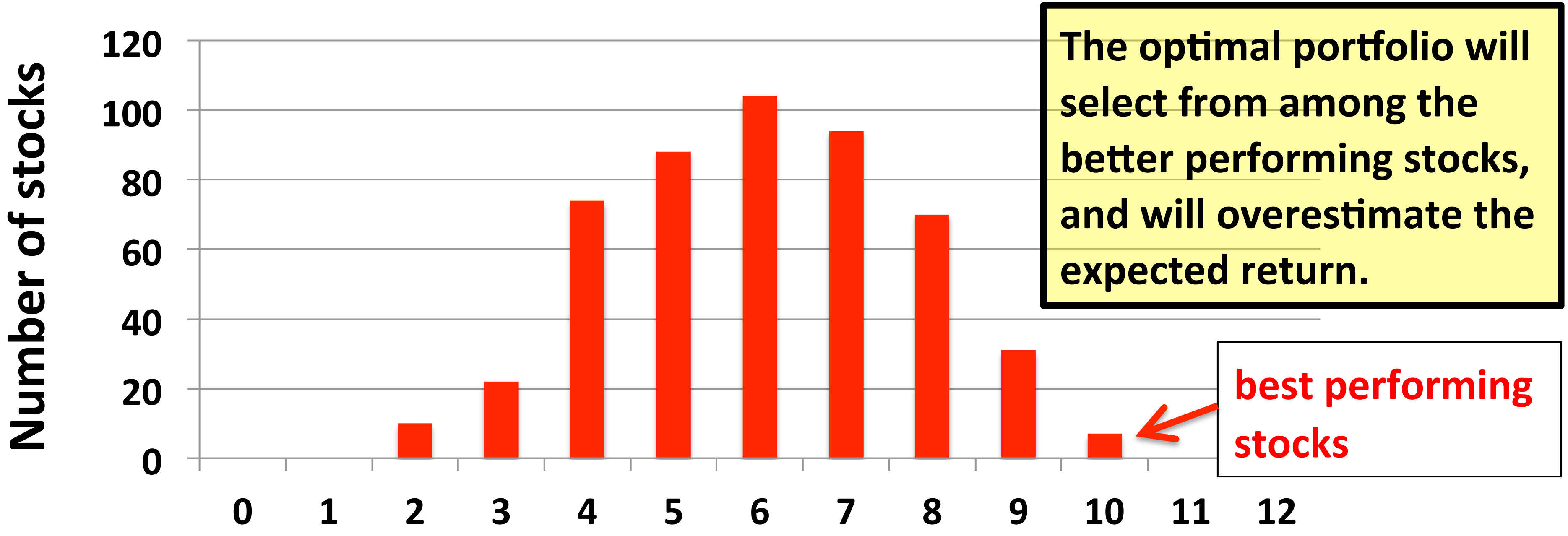
Simulation of the stock market

Suppose we simulated returns from the stock market

- We can toss a coin 12 times to simulate 1 year.
- Heads (tails) on toss j:
  - our stock has a better (worse) return than the S&P 500 in month j.
- Keep track of number of heads and tails.

For any given simulated stock, the expected number of months it outperforms the stock market is 6.

We simulated 500 stocks for one year



Number of months in which a simulated stock beats the average return on the S&P 500.

Past performance

- **Conclusion: if one selects stocks (or mutual funds) based on past performance, one may be selecting stocks that usually perform well, or choosing stocks that were “lucky in the past.”**

- **It is an issue that arises in portfolio optimization.**
- **Dealing with this issue is beyond the scope of this course.**

Final comments on portfolio optimization

- **Variations of this type of model used worldwide.**

- **Special case of CAPM (Capital Asset Pricing Model).**

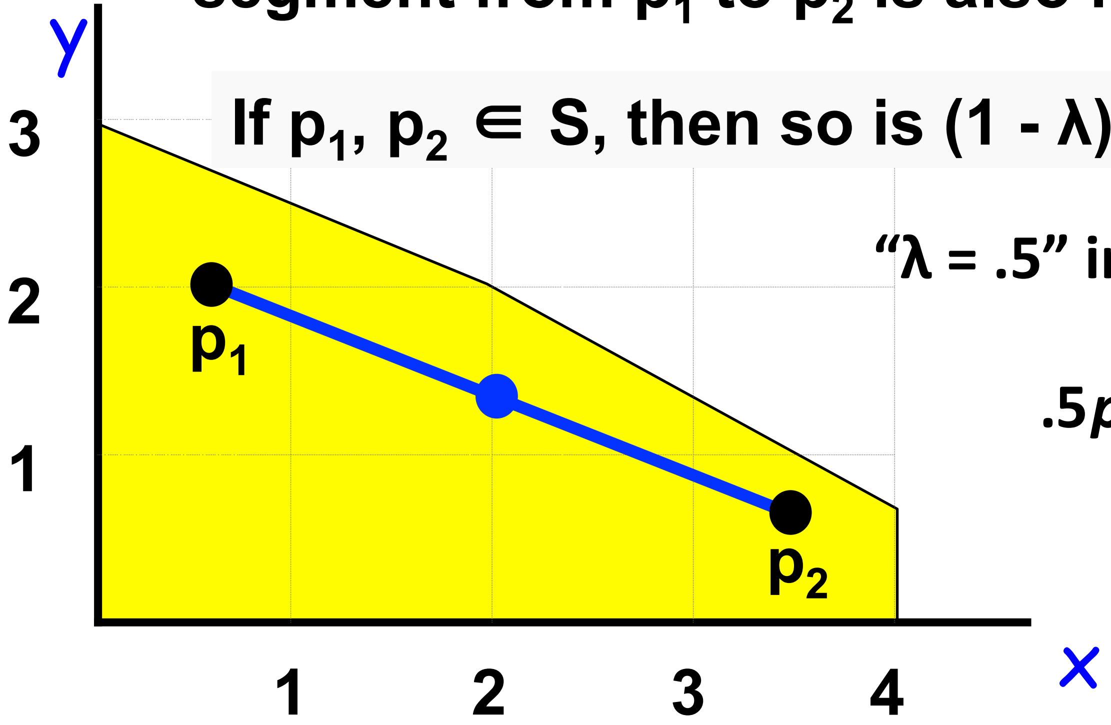
- **Important example of nonlinear programming**

Convex sets and convex functions

Convex Sets

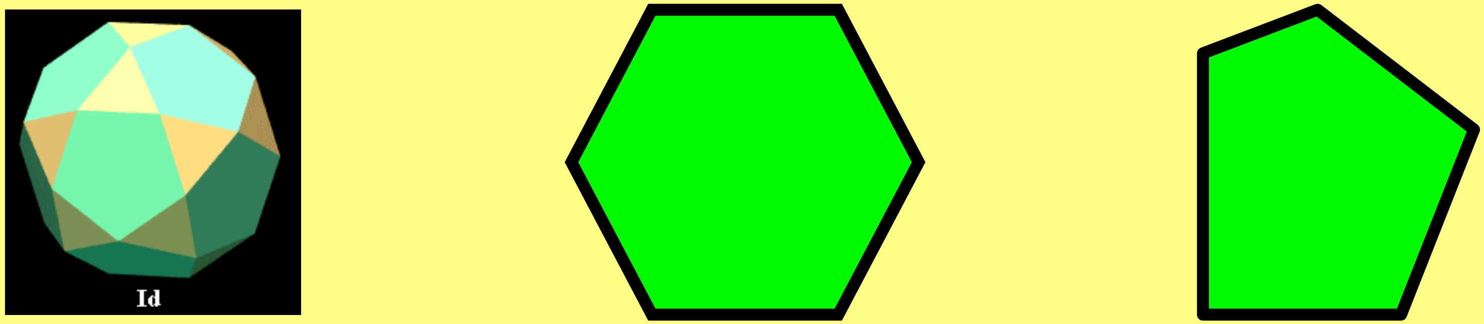
A set S is ***convex*** if for every p_1, $p_2 \in$ S, the line segment from p_1 to p_2 is also in the set;

If p_1, $p_2 \in$ S, then so is $(1 - \lambda)p_1 + \lambda p_2$ for $\lambda \in [0,1]$.

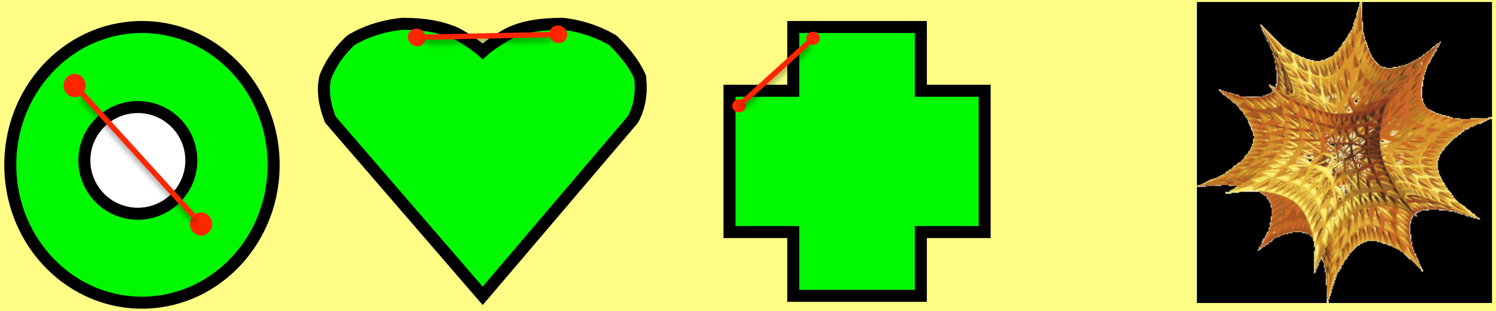


"λ = .5" induces the midpoint.

$$.5p_1 + .5p_2 = \frac{p_1 + p_2}{2}$$



Convex sets



Non-convex sets

Convex functions

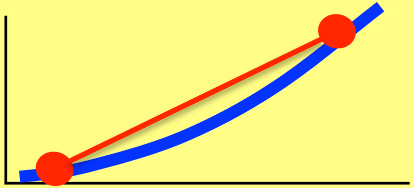
Suppose f is defined on domain D.

A function f is convex if for all x, y ∈ D, and all λ ∈ [0,1]

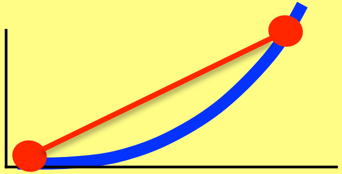
$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

A function f is strictly convex if for all x, y ∈ D, and all λ s.t. 0 < λ < 1

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$$

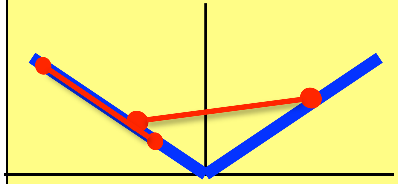


$f(x) = x^2$



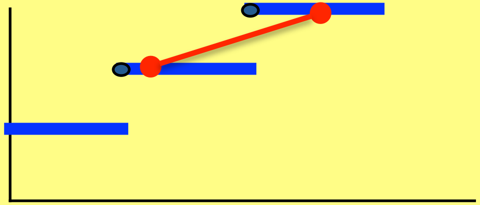
$f(x) = x^3$ for $x \geq 0$

Strictly convex functions

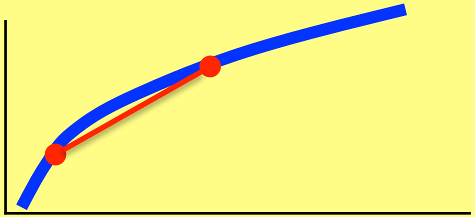


$f(x) = |x|$

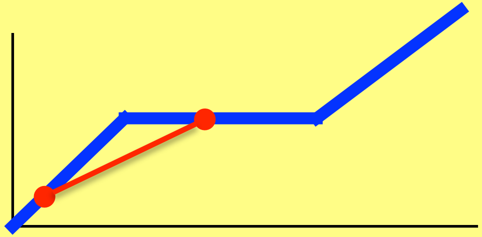
Convex



Step Function



$f(x) = x^{.5}$



whatever

Non-convex functions

More convex functions

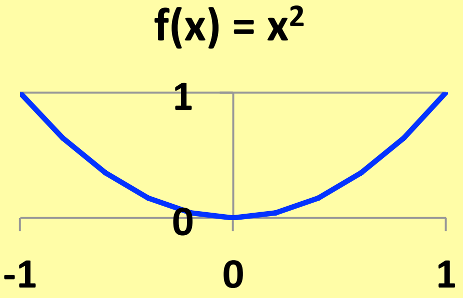
- $f(x) = 1$
- $f(x) = ax + b$ for all a, b
- $f(x) = x^2 + bx + c$ for all b, c
- $f(x) = |x|$
- $f(x) = 1/x$ for $x > 0$
- $f(x) = -\ln(x)$ for $x > 0$
- $f(x) = a^x$ for all $a > 0$

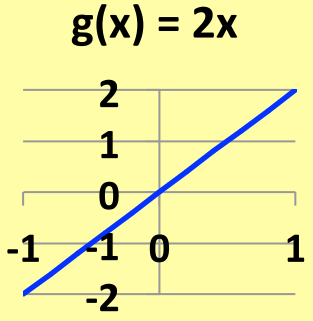
More on convexity

Suppose that f(x) is convex and if g(x) is convex. Then

- f(x) + g(x) is convex
- max {f(x), g(x)} is convex

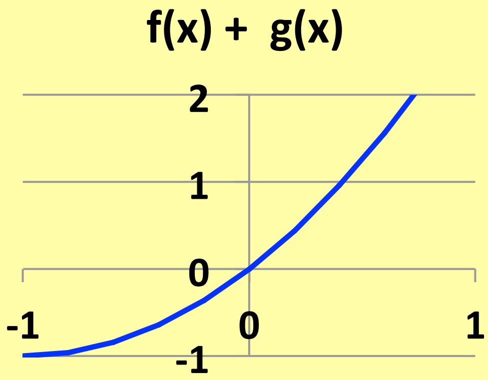
- But min{f(x), g(x)} is possibly non-convex.



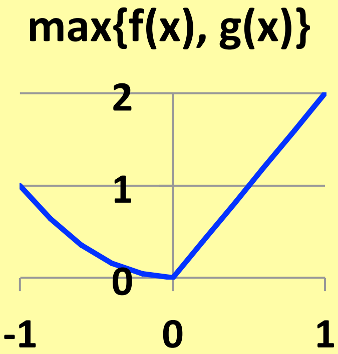


$f(x) = x^2. \quad 0 \leq x \leq 1$

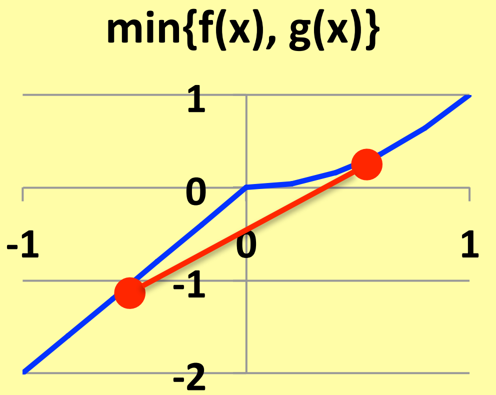
$g(x) = 2x. \quad 0 \leq x \leq 1$



Convex



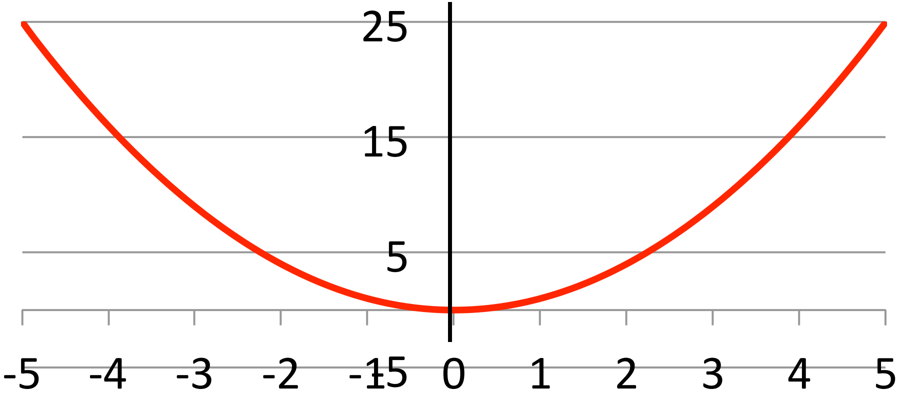
Convex



Not convex

Connection with differentiability

If $f(x)$ is differentiable, and if the second derivative of $f(x)$ is positive for all x, then $f(x)$ is convex.



$$f(x) = x^2$$

$$f'(x) = 2x$$

$$f''(x) = 2$$

Level sets of convex functions are convex sets

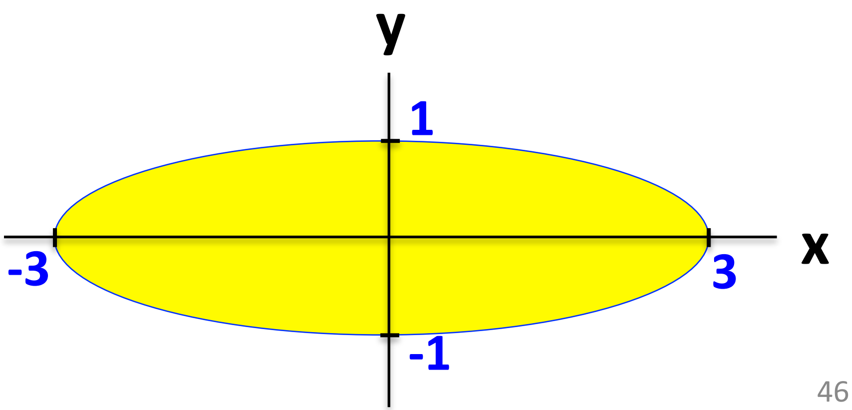
Suppose that f(x) is convex.

Level set of f(x) at α is: $\{x : f(x) \le \alpha\}$.

Theorem. *If f(x) is convex, then each level set is a convex set.*

Example: $f(x, y) = x^2 + 9y^2$

$S(9) = \{(x, y) : f(x, y) \le 9\}$



Local and global optimality for minimization problems

Global minimum for a minimization problem

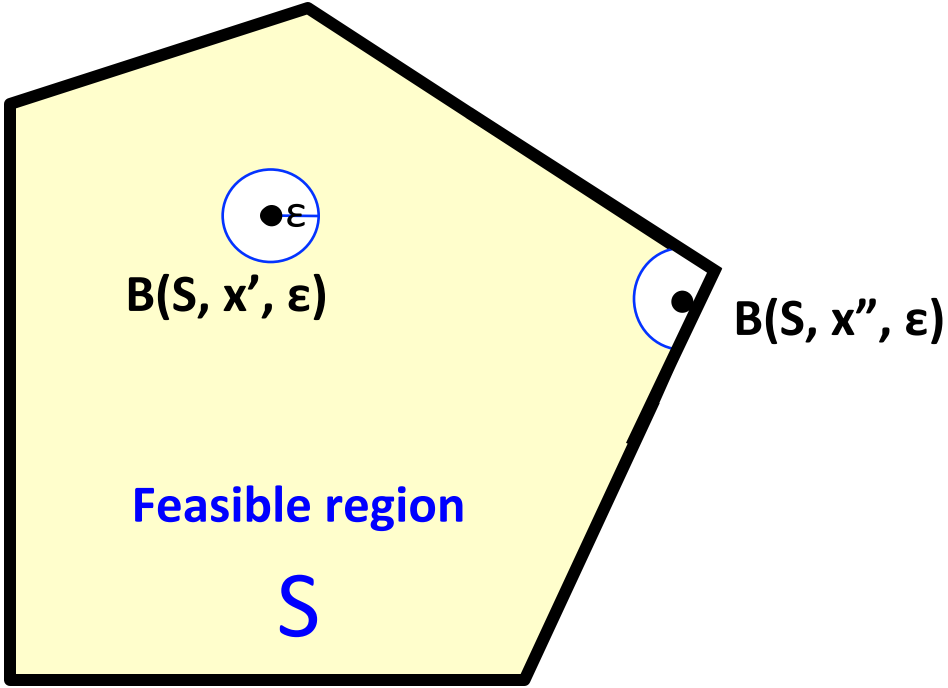
Min	**$f(x)$**	Problem
s.t.	**$x \in S$**	**P**

We say that x' is a global minimum for problem P if

1. **$x' \in S$**
2. **$f(x') \leq f(y)$ for all $y \in S$.**

ε balls in a feasible region

Let x' be a point in S.

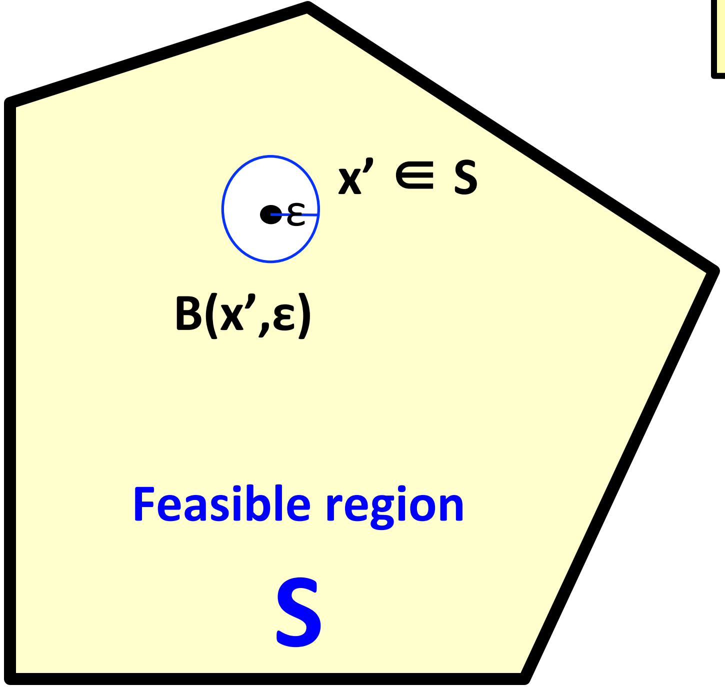


Let B(S, x', ε) be the set of all points in S that are within a distance of ε from x'.

B(S, x', ε) is an ε-ball in S centered at x'.

Local minima

Min	$f(x)$	Problem
s.t.	$x \in S$	P



We say that x’ is a local minimum for problem P if

1. $x' \in S$
2. $f(x') \leq f(y)$ for all $y \in B(S, x', \varepsilon)$ for some $\varepsilon > 0$

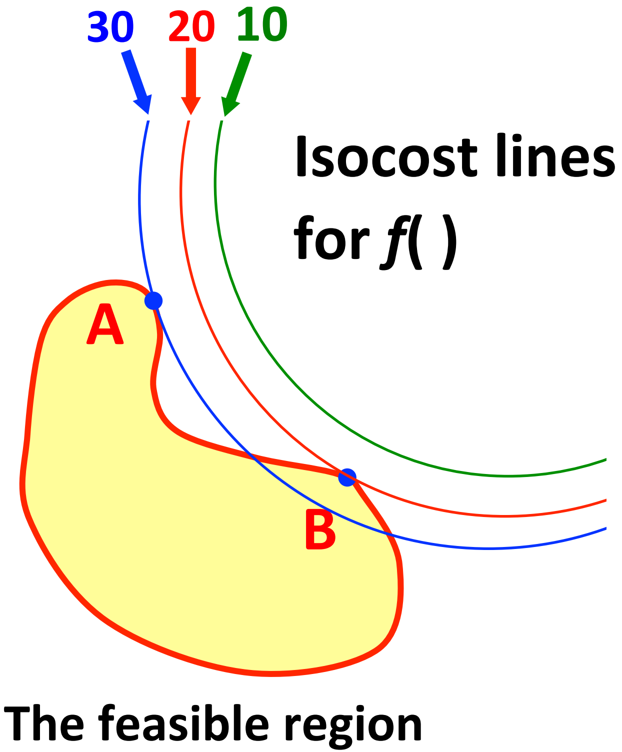
Illustration of global and local minimum

f(A) = 30.

A is a local minimum

f(B) = 20.

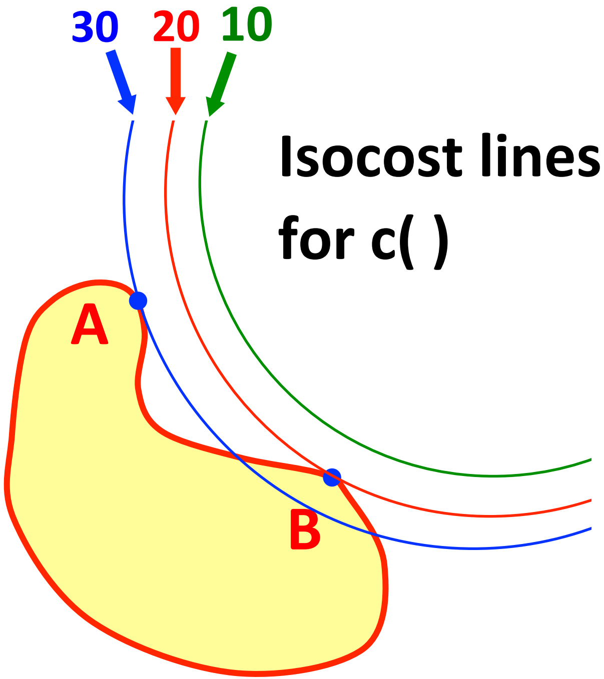
B is the global minimum



Nonlinear programming algorithms and local minima

Nonlinear programming algorithms generally stop when they find a local minimum.

Best minimization problems: those for which a local minimum is guaranteed to be a global minimum.



The feasible region

Convex minimization problems

A nonlinear programming minimization problem is called a convex minimization problem if the objective function is convex and the feasible region is a convex set.

Theorem. ***If x is a local minimum for a convex minimization problem, then x is also a global minimum.***

That is, convex minimization problems are "easy to solve."

Concave maximization problems

Recall: A function f( ) is concave if –f( ) is convex.

A nonlinear programming maximization problem is called a concave maximization problem if the objective function *f*( ) is concave and the feasible region is a convex set.

Theorem. ***If x is a local maximum for a concave maximization problem, then x is also a global maximum.***

Two applications of nonlinear programming

When prices depend on supplies

MIT Corp sells up to 1,000 red widgets for a profit of \$3.50 per widget. It can sell more. For every δ red additional widgets sold, the profit (for each of the 1,000 + δ red widgets) is reduced by $.002 \times \delta$. How many should be sold, assuming no other constraints?

The amount sold is at least 1000.

$$\text{Profit}(\delta) = \underbrace{\left(3.50 - .002\,\delta\right)}_{\text{price}} \times \underbrace{\left(1{,}000 + \delta\right)}_{\text{quantity}}$$

$$= -.002\,\delta^2 + 1.5\,\delta + 3{,}500$$

$$\frac{d(\text{Profit})}{d\delta} = -.004\,\delta + 1.5$$

To find the maximum, try setting the derivative to 0.

$-.004\ \delta + 1.5 = 0 \Rightarrow \delta = 375.$

Profit = $3,781.25

If price depends on the quantity sold,

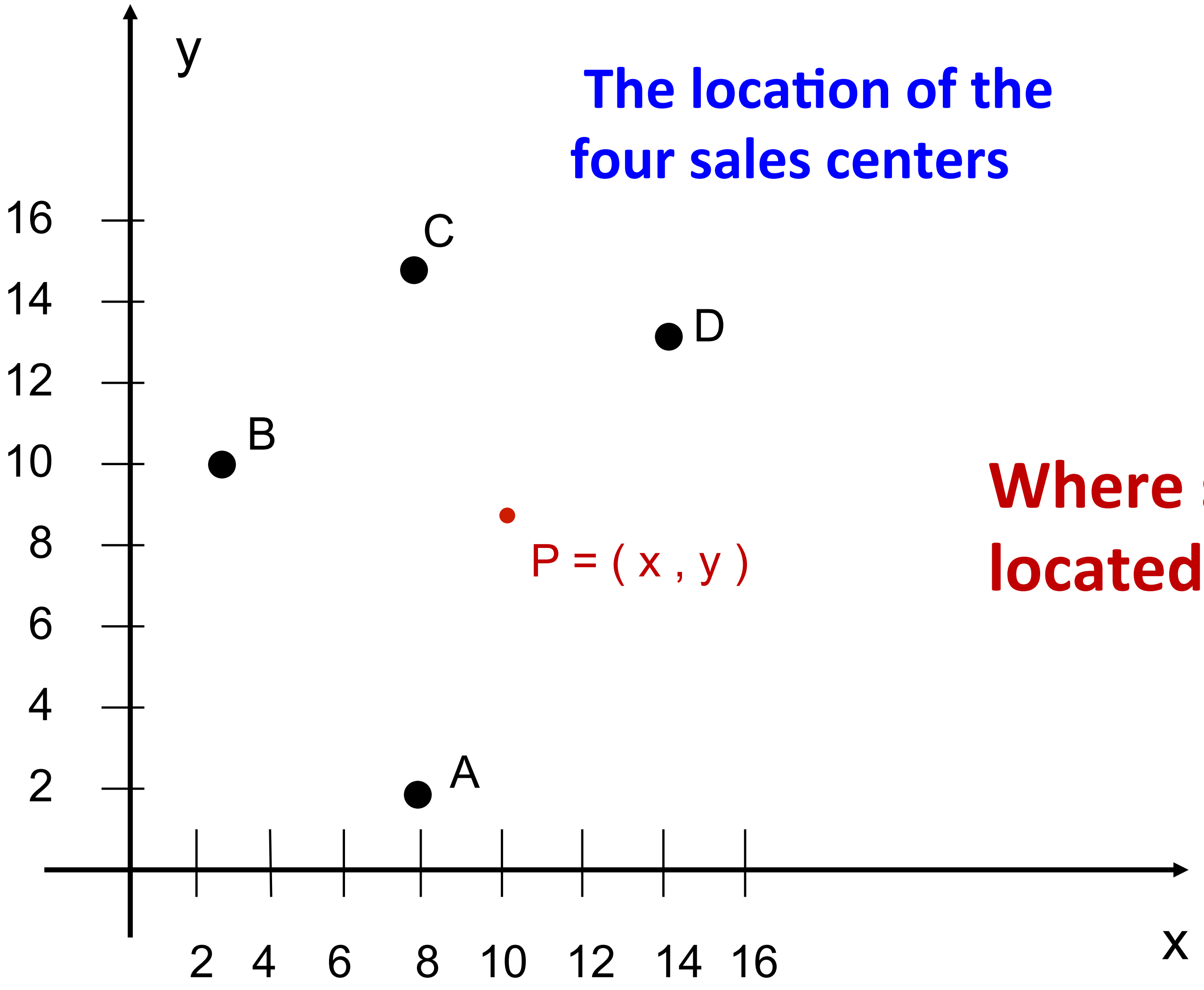
then the profit is a nonlinear function.

Optimal Facility Location

Locate a new warehouse to minimize total travel costs to four sales centers. Travel cost = $1 per mile.

Sales Center	Code/Coordinates		Daily Truck Deliveries
Amherst	A	(8,2)	9
Boston	B	(3,10)	7
Colby	C	(8,15)	2
Dartmouth	D	(14,13)	5

Where should the firm locate the new warehouse?



The location of the four sales centers

Where should P be located?

Let P = (x , y) be the location of the warehouse.

The distance from P to A = (8, 2) is $d(P,A)=\sqrt{\left(x-8\right)^2+\left(y-2\right)^2}$

d(P, B), d(P, C), d(P, D) are defined similarly.

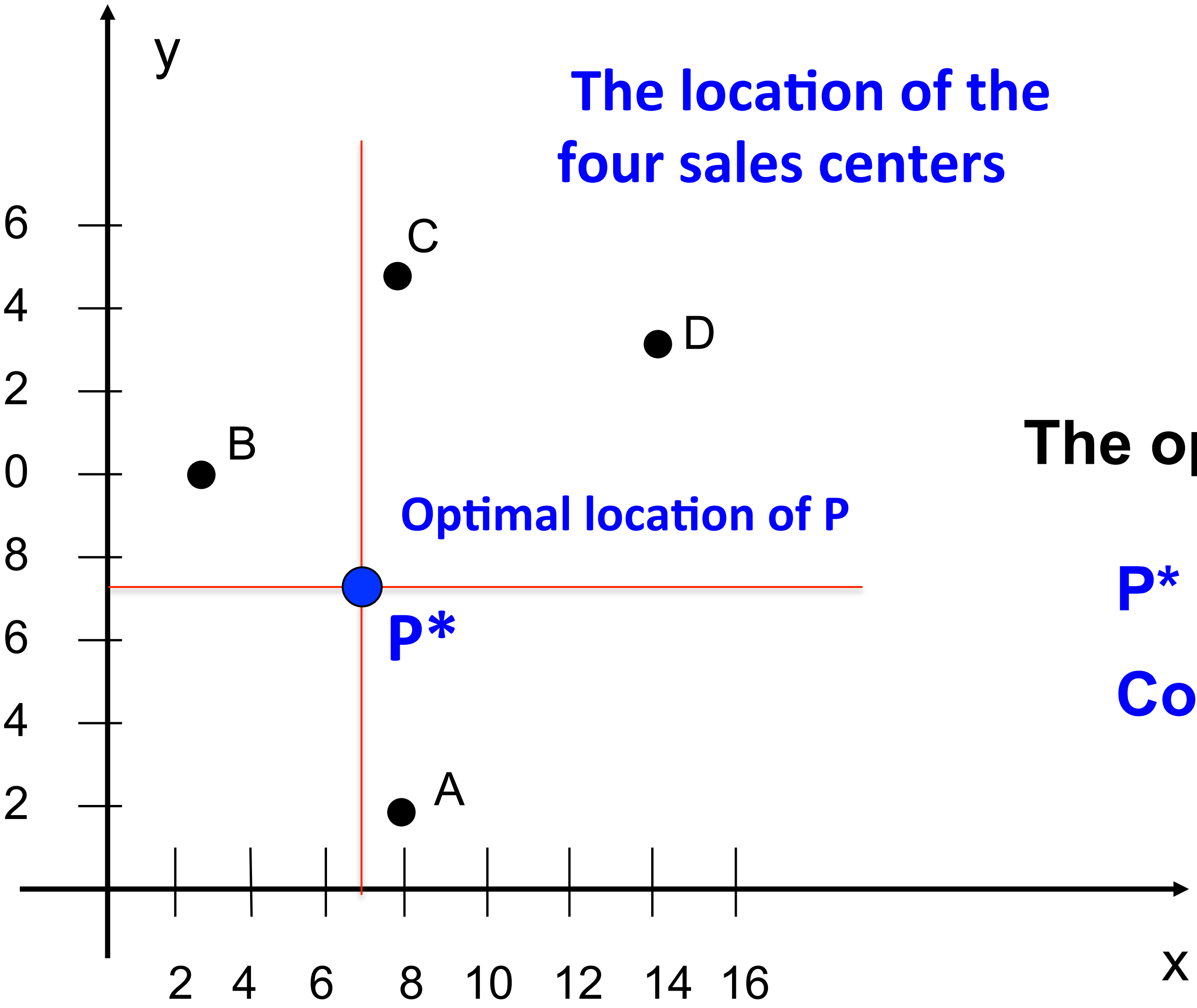
Min [9 d(P, A)] + [7 d(P, B)]
+ [2 d(P, C)] + [5 d(P, D)]

s.t. P = (x, y)

The optimal solution is :

P* = (6.95 , 7.47)

Cost = $142.97

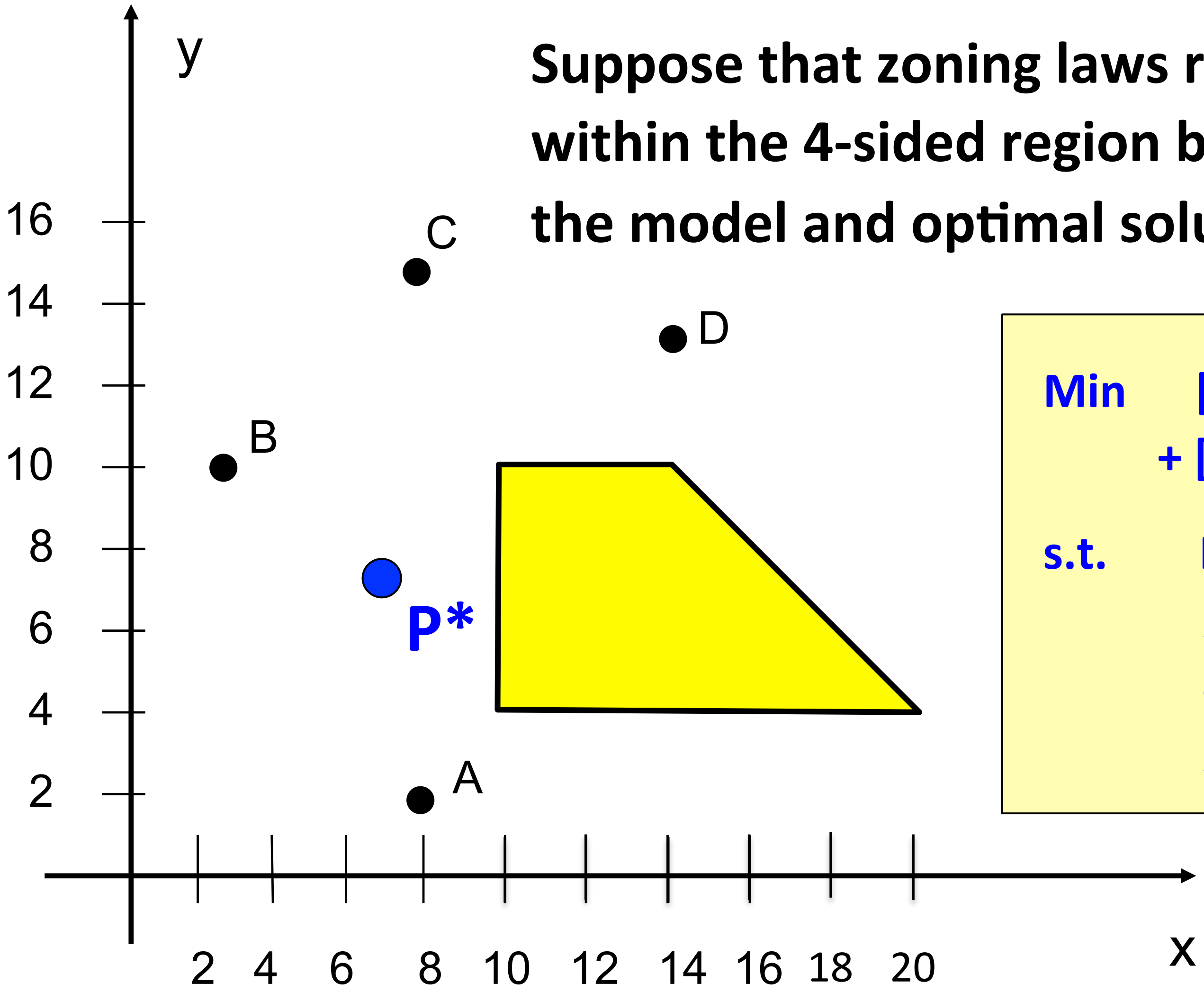


The optimal solution is :

P* = (6.95 , 7.47)

Cost = $142.97

Suppose that zoning laws required it to be within the 4-sided region below. How do the model and optimal solution change?



Min $[9\ d(P, A)] + [7\ d(P, B)] + [2\ d(P, C)] + [5\ d(P, D)]$

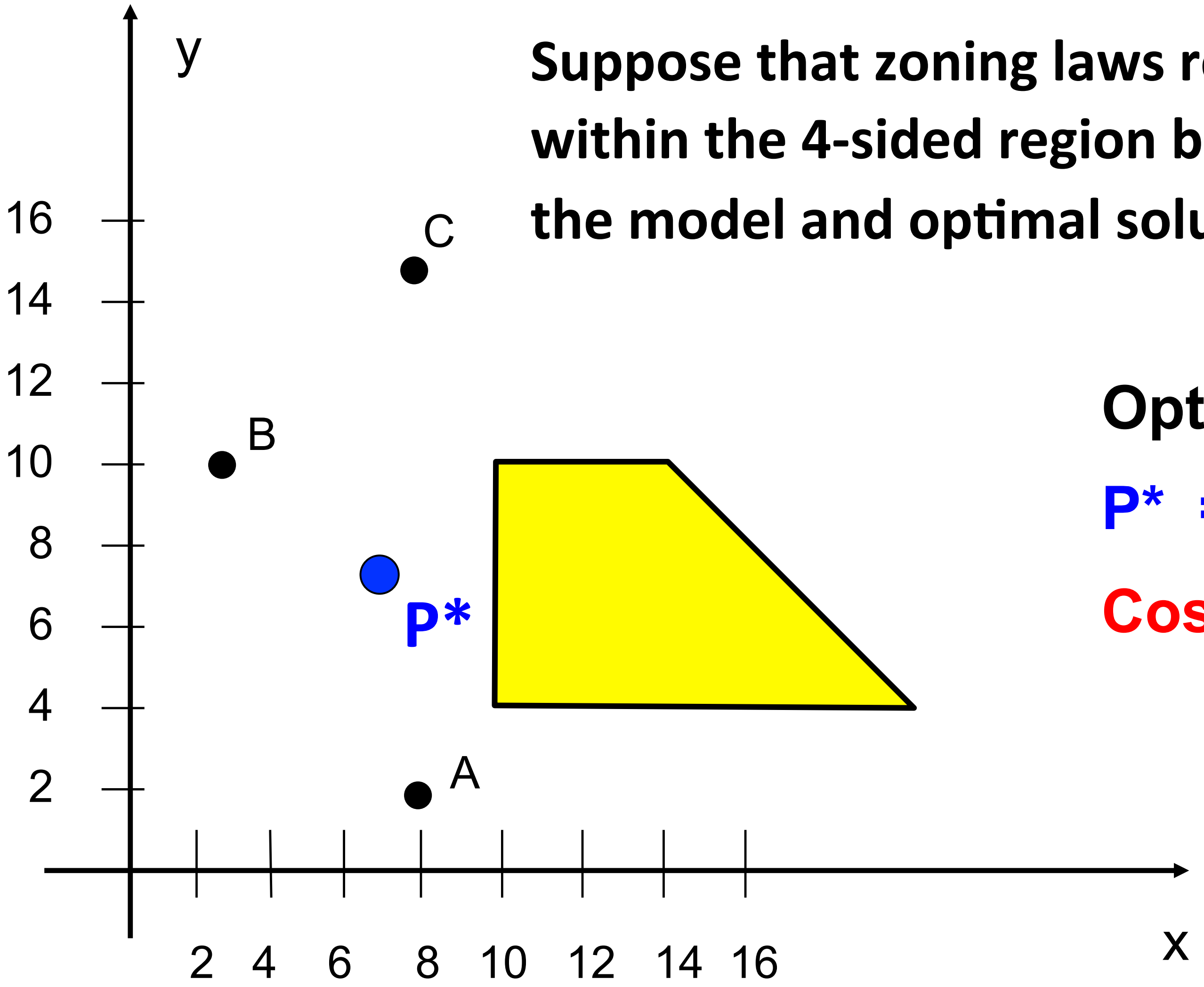
s.t. $P = (x, y)$

$x \geq 10$

$4 \leq y \leq 10$

$x + y \leq 24$

Suppose that zoning laws required it to be within the 4-sided region below. How do the model and optimal solution change?



Optimal is :

P* = (10.0, 7.4)

Cost = $154.22

Cost has increased by $11.25.

The constrained location problem

Before solving the problem.

Decision variables

P = (x, y)	x	y
	0.000	0.000

The four points			**Distance from P**	**Weights (costs)**	**Weighted distance**
A	8	2	8.246	9	74.216
B	3	10	10.440	7	73.082
C	8	15	17.000	2	34.000
D	14	13	19.105	5	95.525

Total weighted distance 276.823

Constraints

x >= 10	0.000	>=	10
y >= 4	0.000	>=	4
y <= 10	0.000	<=	10
x+y <= 24	0.000	<=	24

The constrained location problem

After solving the problem.

Decision variables

P = (x, y)	x	y
	10.000	7.400

The four points			**Distance from P**	**Weights (costs)**	**Weighted distance**
A	8	2	5.759	9	51.829
B	3	10	7.467	7	52.270
C	8	15	7.858	2	15.717
D	14	13	6.882	5	34.408

Total weighted distance 154.224

Constraints

x >= 10	10.000	>=	10
y >= 4	7.400	>=	4
y <= 10	7.400	<=	10
x+y <= 24	17.400	<=	24

2^{nd} Order Cone Programming:

An easy convex minimization problem

Euclidean norms

Notation: Suppose that $y = y_1, y_2, \ldots, y_k$

y is a vector of k decision variables.

Then $$\|y\|_2 = \left(\sum_{i=1}^{k} (y_i)^2 \right)^{1/2}.$$

We refer to $\|y\|_2$ as the Euclidean norm of vector y.

Second order cone program (SOCP)

Linear program

- Linear objective
- Linear inequalities and equalities.

SOCP

- same as LP except that it also includes constraints such as: "Euclidean norm + linear constraint ≤ RHS"**

e.g., $\|y\|_2 + 3x_1 - x_5 \leq 10$

** A slightly more general form is permitted.

Second order cone program (SOCP)

- Lots of applications
- Special case of convex minimization (or concave maximization)
- Almost as easy to solve as linear programs
- Lots of available software
  - CPLEX
  - Gurobi

Example 1: portfolio optimization for independent random variables

$$\max \quad E(W) = 65\,I + 80\,P + 50\,R$$

$$\text{s.t.} \quad SD(W) \le 12$$

$$I + P + R = 1$$

$$I \ge 0, \quad P \ge 0, \quad R \ge 0$$

"12" is the target value.

We would optimize for multiple target values.

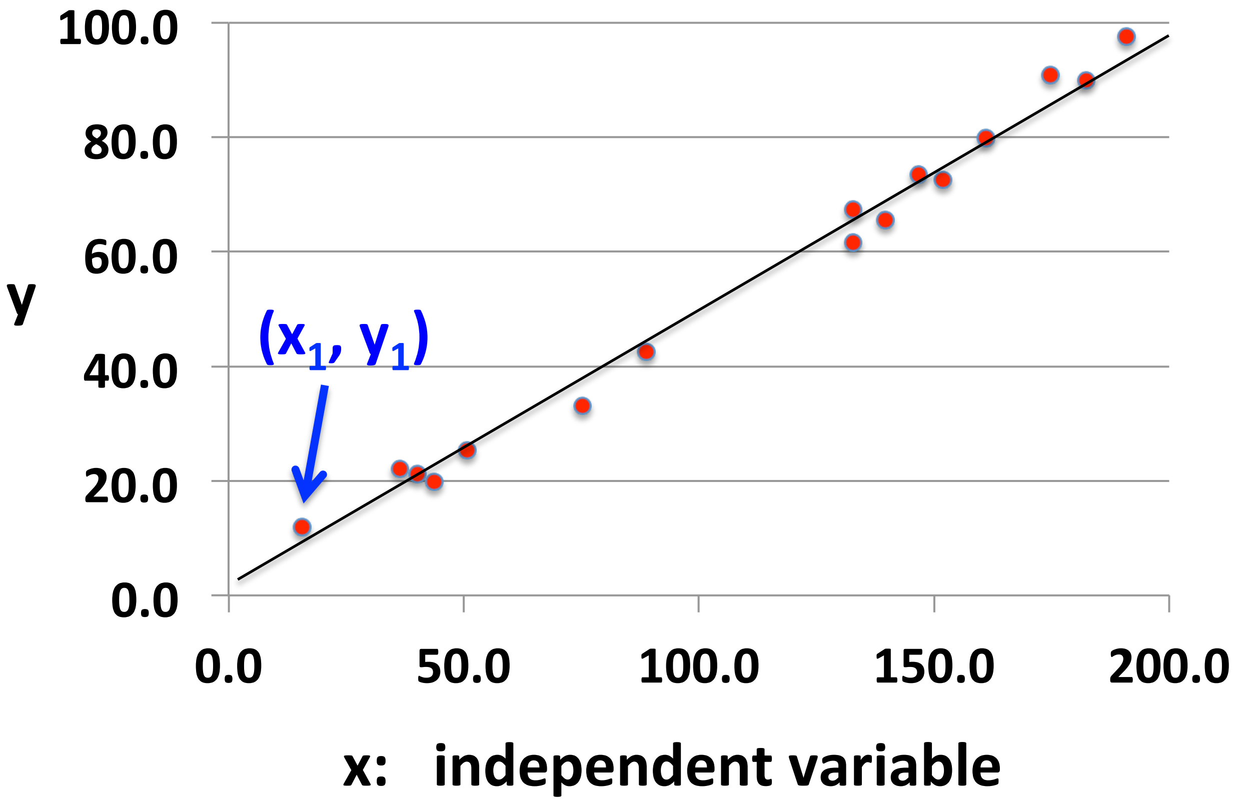
Getting it into the right form

$$SD(W) = \left(225\, I^2 + 324\, P^2 + 100\, R^2 \right)^{.5} \leq 12$$

$y_1 = 15\, I$; $y_2 = 18\, P$; $y_3 = 10\, R$

$$\|y\|_2 \leq 12$$

Example 2: Simple linear regression



Estimate y as a linear function of x.

Choose β_0 and β_1 and estimate y as follows: for all *j*:

$$\hat{y}_j = \beta_0 + \beta_1 x_j$$

$$\boxed{\min \sum_{j=1}^{N} (y_j - \hat{y}_j)^2}$$

Getting it into the right form

$$\boxed{\min \sum_{j=1}^{N} (y_j - \hat{y}_j)^2}$$

We need to move the quadratic term to the constraints.

$$w_j = (y_j - \hat{y}_j) \qquad \text{for all } j.$$

$$\left\| w \right\|_2 \leq T$$

Iteratively search for the least value of T.

SOCP solves much faster

- If you have a convex minimization problem that can be reformulated as a SOCP, then do so.
- You will be glad you did.

Guidelines for solving nonlinear problems

Guidelines in Solving Nonlinear Problems

Guideline 1: Unlike linear optimization, nonlinear optimization may be difficult to solve even with today's computers.

Guideline 2: Algorithms for solving nonlinear optimization rely on calculus. They are only guaranteed to find a locally optimal solution, which is not necessarily a globally optimal solution.

Guidelines in Solving Nonlinear Problems

Guideline 3: Some nonlinear optimization problems are easy to solve, others are difficult. The difficulty depends on the model's mathematical structure. The easiest to solve are convex minimization problems (and concave maximization problems). and the easiest of these are SOCPs and LPs.

Guideline 4: Software for solving nonlinear optimization models varies with the degree of functionality and with the price. The better software packages solve nonlinear models with many variables and/or constraints.