Optimization – Part 1

P. Agius – L2.1, Spring 2008

Contents

What is Optimization?

“ … arrive at the best possible decision in any given set of circumstances.” [G.R. Walsh, *Methods of Optimization*]

“Optimization models attempt to express, in mathematical terms, the goal of solving a problem in the ‘best’ way.”

[Nash and Sofer, *Linear and Nonlinear Programming*]

Notation

$$\begin{array}{ll} \min_x & \mathrm{f}(\mathrm{x}) \longleftarrow \text{Objective function} \\ \text{subject to} & \left.\begin{array}{l} g(x) \geq 0 \\ h(x) = 0 \end{array}\right\} \text{Constraints} \\ & x \in R \end{array}$$

$$\min_x f(x) \text{ is equivalent to } \max_x -f(x)$$

In **unconstrained optimization**, there are no constraints.

$$\begin{aligned} \min_x \quad & \mathrm{f(x)} \\ \text{subject to} \quad & g(x) \geq 0 \\ & h(x) = 0 \\ & x \in R \end{aligned}$$

←Inequality constraints

←Equality constraints

Linear constraints (Eg. $x + 2 = 5$, $x_1 - x_2 > 6$)
Nonlinear constraints (Eg. $x_1^2 + x_2^2 \geq 5$)

A function f is convex on a set S if

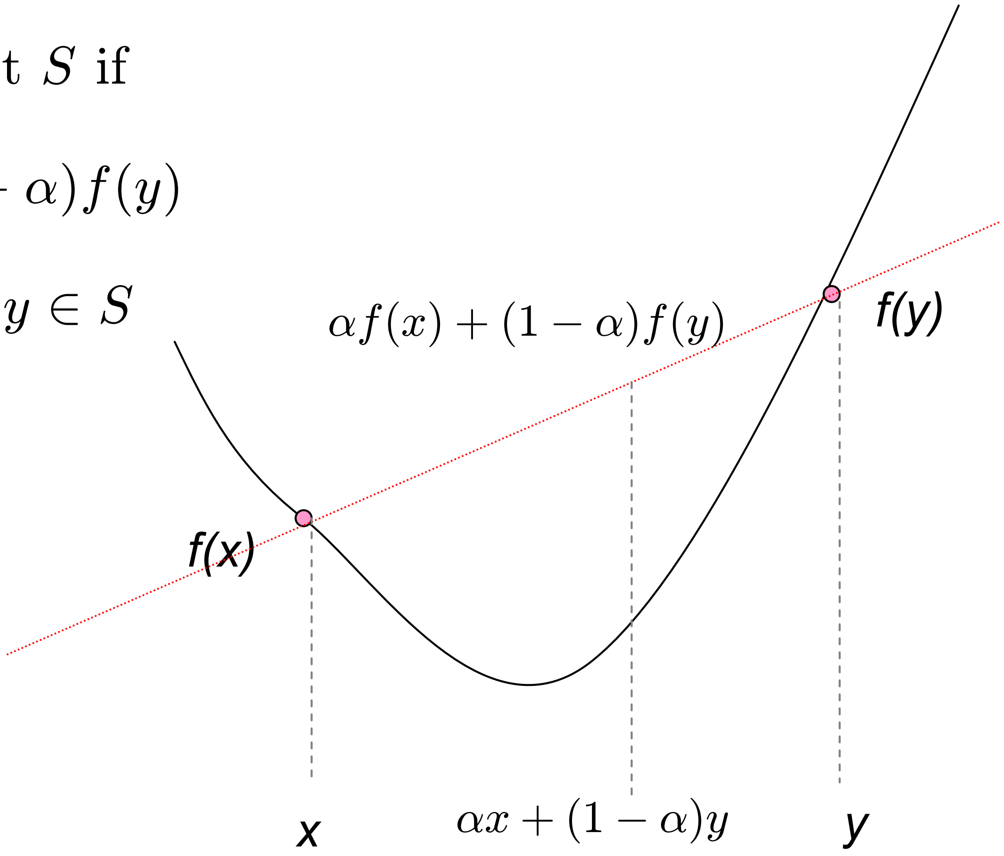
$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$

for all $0 \leq \alpha \leq 1$ and for all $x, y \in S$

A function f is convex on a set S if

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

for all $0 \leq \alpha \leq 1$ and for all $x, y \in S$



Convex quadratic function … more later

$$\min_x \quad 0.5\mathrm{x}^T Q x$$

Formulating the Problem

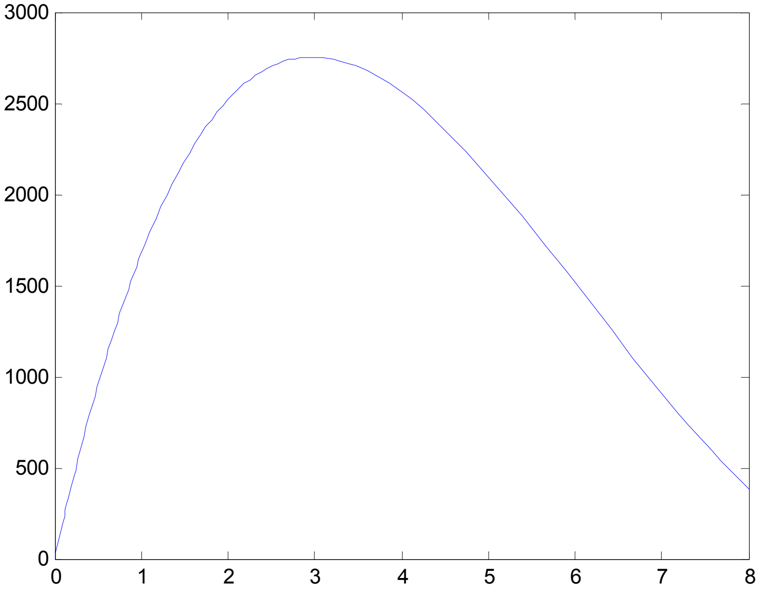
Example: Find two nonnegative numbers whose sum is 9 such that the product of one number and the square of the other number is a maximum.

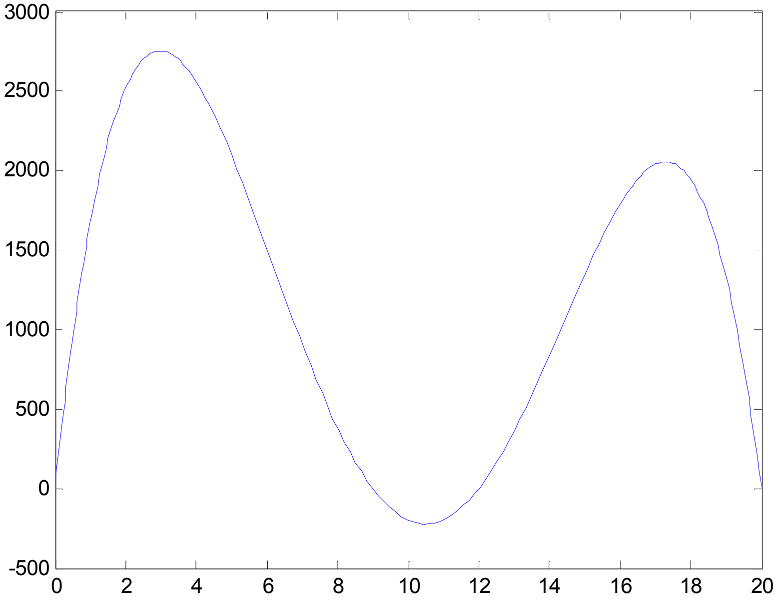
Let numbers be x and $9 - x$.

$$\begin{array}{ll} \max_x & \mathrm{x(9\text{-}x)}^2 \\ \text{subject to} & x \geq 0 \end{array}$$

Unimodal functions

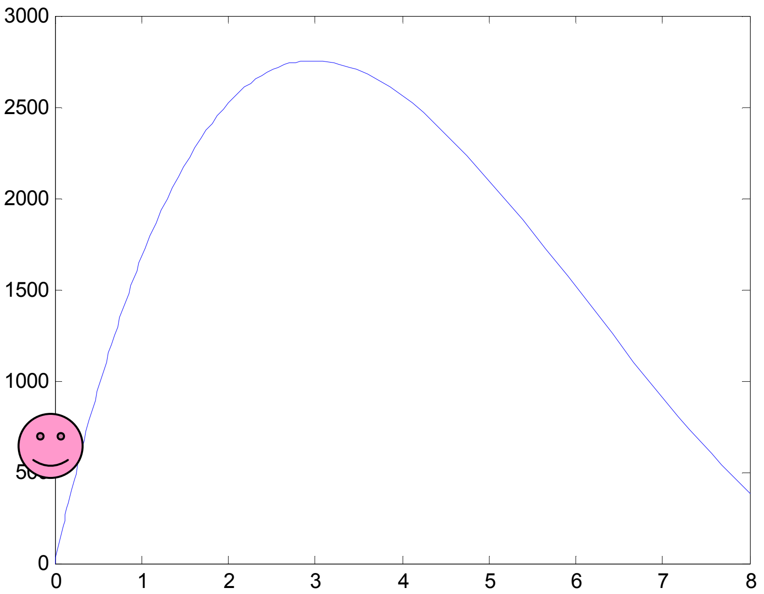
A function is **unimodal** if it has only one maximum or minimum.

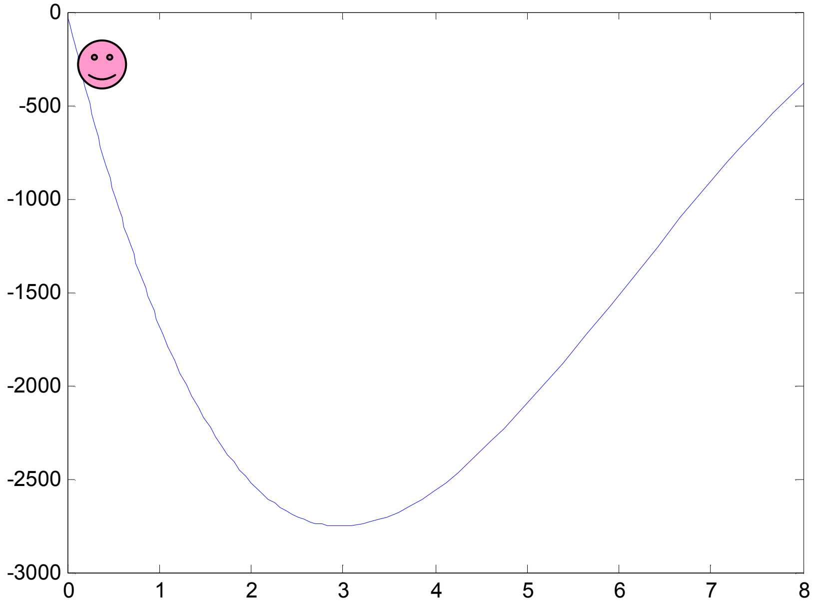




Stationary points … $\nabla f\,(x^*) = 0$

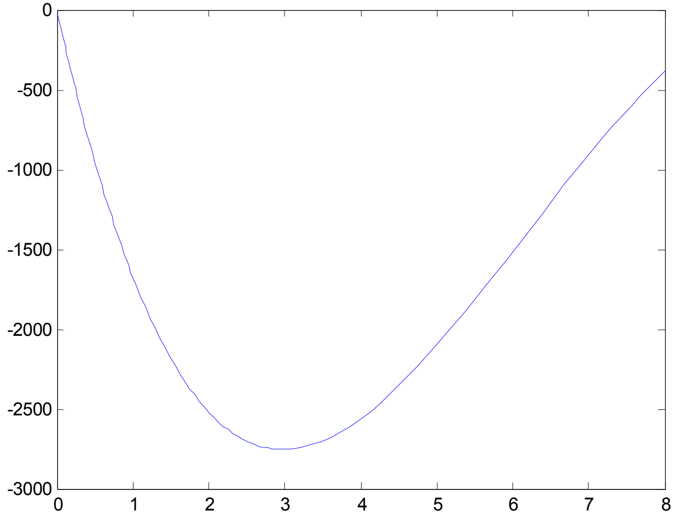
For unconstrained problems, a stationary point is a point where the first derivatives of f are equal to zero.



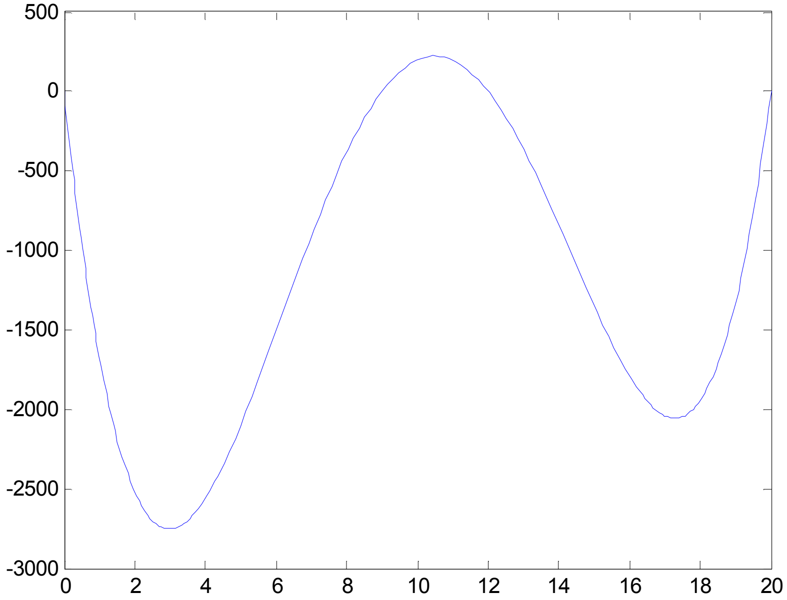


Local and Global minimizers

A **local** minimizer is a point x^* that satisfies the condition $f(x^*) \leq f(x)$ for all x in a search region.



A **global** minimizer is a point x^* that satisfies the condition $f(x^*) \leq f(x)$ for all x.



STRICT local/global → $f(x^*) < f(x)$

Unconstrained Optimization - FONC

A [strict] local minimizer is a point x^* that satisfies the condition $f(x^*) \le f(x)$ $[f(x^*) < f(x)]$ for all x.

Example: $f(x) = x^2$ $x^* = 0$

If x^* is a local minimizer and satisfies the condition

$$\boxed{\nabla f(x^*) = 0}$$

∇ = first derivative

This is known as the First Order Necessary Condition

$\nabla f(x) = 2x$, $\nabla f(x^*) = 0$.

Back to the previous example … (constrained example)

Let numbers be x and $9 - x$.

$$\begin{array}{ll} \max_x & \text{x(9-x)}^2 \\ \text{subject to} & x \geq 0 \end{array}$$

Set the first derivative to zero:

$$-x^2 + 12x - 27 = 0$$

Solutions: $x = 3$ or $x = 9$

Unconstrained Optimization - SONC

For **one-dimensional problems**,
define $\nabla^2 f\ (x^*)$ (the second derivative of f wrt x)
to be **p**ositive **s**emi-**d**efinite (psd) if

$$\boxed{\nabla^2 f\ (x^*) \geq 0}$$

(**p**ositive **d**efinite if $\nabla^2 f\ (x^*) > 0$)

Example: $f(x) = x^2 \quad x^* = 0$
$\nabla^2 f\ (x) = x \quad \nabla^2 f\ (x^*) = 0$

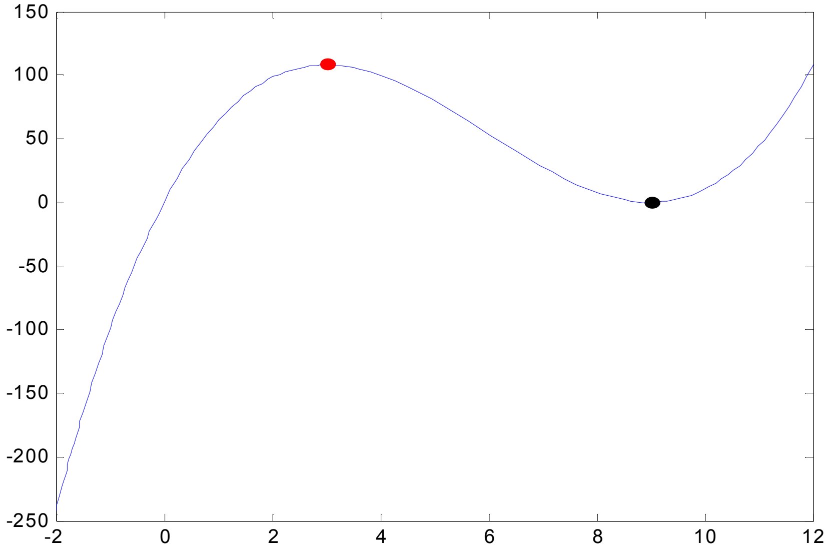
If x^* is a local minimizer
then $\nabla^2 f\ (x^*)$ is psd –
this is the **S**econd **O**rder **N**ecessary **C**ondition.

Example continued….

Set the first derivative to zero:

$$-x^2 + 12x - 27 = 0$$

Solutions: $x = 3$ or $x = 9$



Second derivative:

$$\nabla^2 f(x) = -2x + 12$$

✓ $$\nabla^2 f(3) = -2 * 3 + 12 > 0$$

$$\nabla^2 f(9) = -2 * 9 + 12 < 0$$

Unconstrained Optimization - SOSC

If $\nabla f\,(x^*)=0$ and $\nabla^2 f\,(x^*)$ is positive definite then x^* is a strict local minimizer of f.

Example: $f(x) = x^2$ $x^* = 0$
$\nabla^2 f\,(x) = x$ $\nabla^2 f\,(x^*) = 0$

This is the **S**econd **O**rder **S**ufficiency **C**ondition.

A two-dimensional example for SOSC

Need a little bit of matrix notation and definitions.

Matrix notation
http://www.purplemath.com/modules/matrices2.htm
Matrix operations
http://math.jct.ac.il/~naiman/linalg/lay/slides/c02/sec2_1ov.pdf
Vector dot products (from Wikipedia)
http://en.wikipedia.org/wiki/Dot_product

Some notation that I will use: [a b c d]' is the *transpose of the vector* containing elements a,b, c and d. (on occasion, 'T' will be used to indicate the transpose). In matrices, ';' indicates a new line of elements.

Classwork
Q1: If v1=[5 4 7 1] and v2=[3 1 2 9], find <v1,v2>, ||v1|| and ||v2||.
Q2: Find the product of the following two matrices:
[1 1 0; 2 1 5; 3 2 2] and [3 1 1; -2 0 8; -1 -3 0]

A two-dimensional example for SOSC

Consider the function

$$f(x_1, x_2) = \frac{1}{3}x_1^3 + \frac{1}{2}x_1^2 + 2x_1x_2 + \frac{1}{2}x_2^2 - x_2 + 9$$

Stationary point: $\bigtriangledown f(x) = \begin{pmatrix} x_1^2 + x_1 + 2x_2 \\ 2x_1 + x_2 - 1 \end{pmatrix} = 0$

Derivative w.r.t. x_1

Derivative w.r.t. x_2

Substitute second component $x_2 = 1 - 2x_1$ into first component, we get two solutions: $x_a = [1 - 1]$ and $x_b = [2 - 3]$.

SOSC example continued …

Second derivative, also known as the Hessian …

Derivative of 1st component of ∇f wrt x_1

Derivative of 1st component of ∇f wrt x_2

$$\nabla^2 f(x) = \begin{pmatrix} 2x_1 + 1 & 2 \\ 2 & 1 \end{pmatrix}$$

Derivative of 2nd component of ∇f wrt x_1

Derivative of 2nd component of ∇f wrt x_2

$$\text{Is } \nabla^2 f(x_a) \text{ psd? Is } \nabla^2 f(x_b) \text{ psd?}$$

SOSC example continued ...

A matrix *H* is said to be **positive semi-definite** if either of the following is true:

- x' H x ≥ 0 for all vectors x
- all of the eigenvalues of H are ≥ 0

More on eigenvalues later on ...

$$\nabla^2 f(x_a) = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix}$$

Try x=[-1 1] ...

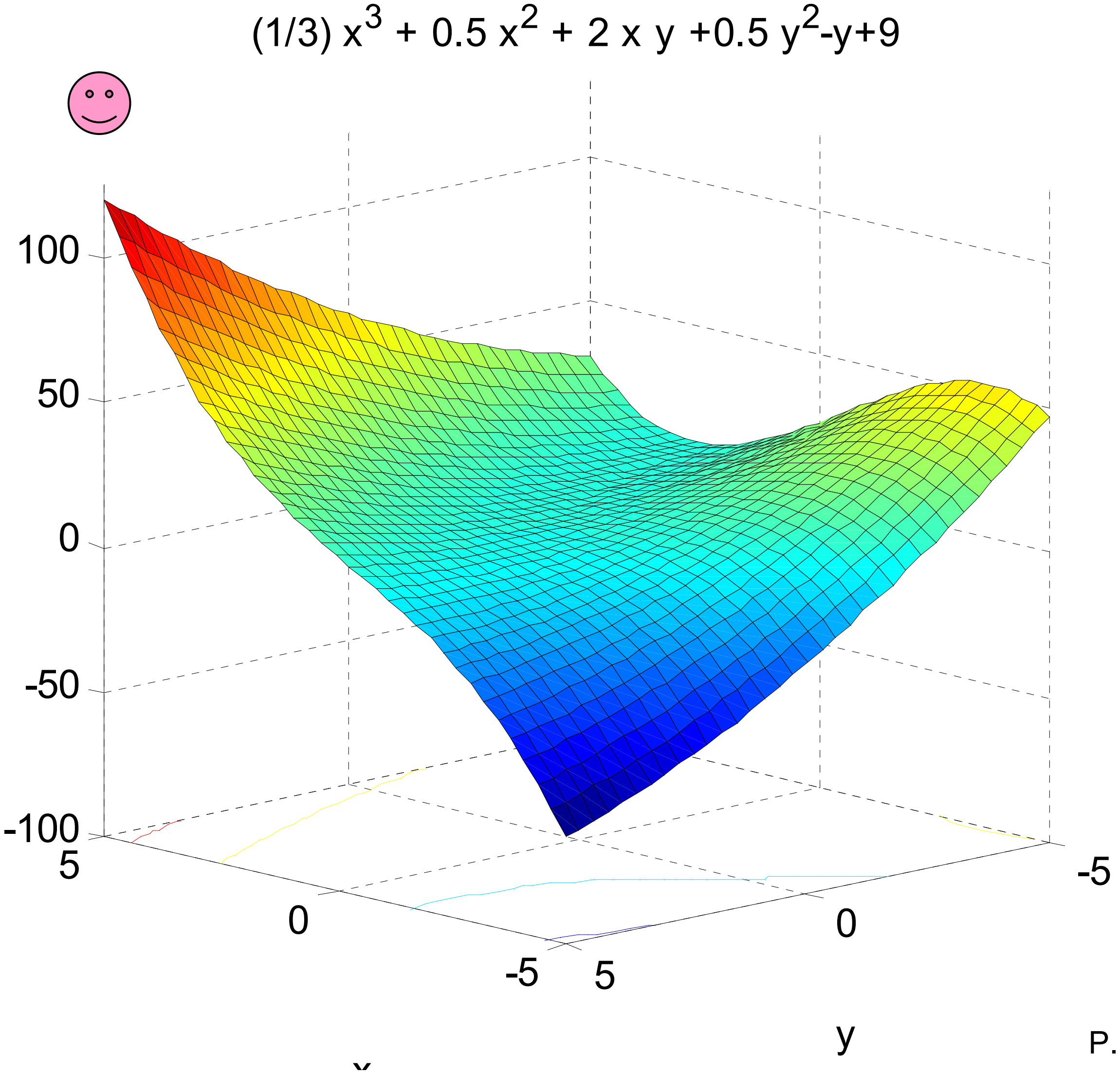
... get 0. Not positive definite.
SOSC failed.

$$\nabla^2 f(x_b) = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$$

This is positive definite.
So x_b is a local minimizer.

In matlab …

`ezsurfc('(1/3)*x^3 + 0.5*x^2 + 2*x*y +0.5*y^2-y+9',[-5,5,-5,5],35)`

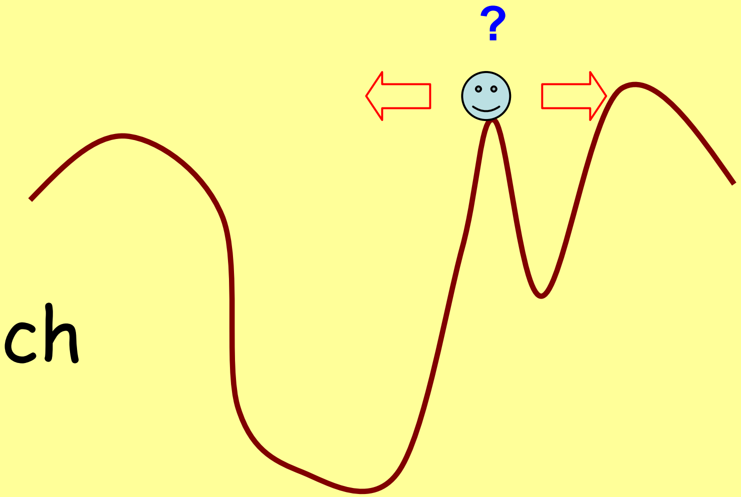


Searching for x*

Let f(x) be unimodal. We wish to find x* by function evaluations and comparisons.

- how should we choose values of x that bring us as close as possible to x*?
- can we choose those values such that the number of iterations is minimal?

- Which direction (*p*)?
- How big a step (*a*)?
  - Steepest descent
  - Golden Section Search
  - Newton's Method



Steepest descent (*p*)

At the k^{th} iteration, the direction is defined to be

$$p_k = - \bigtriangledown f(x_k)$$

and then uses a line search to find

$$x_{k+1} = x_k + \alpha_k p_k$$

where α_k is the step-size

Exact line search (α)

Choose $\alpha_k > 0$ that is a local minimizer of $F(\alpha) = f(x_k + \alpha p_k)$.

Note that $\bigtriangledown F(\alpha) = p_k \bigtriangledown f(x_k + \alpha p_k)$.

Example: Consider the function $f(x) = x^2 - 6x + 5$ and suppose $x_k = 4$.

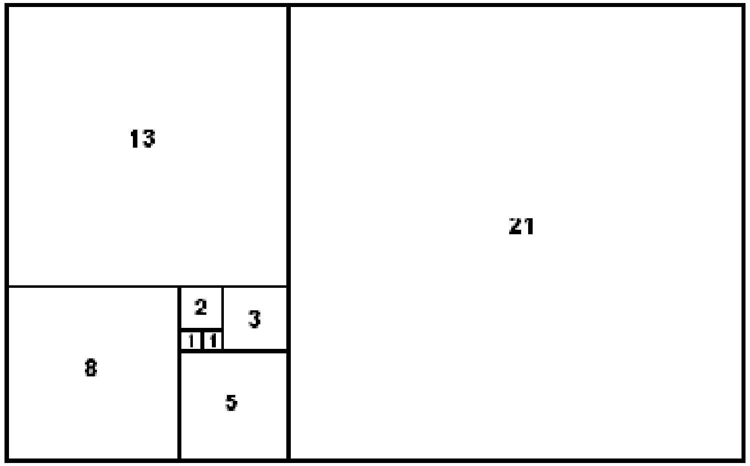
Using Steepest descent, $p_k =$???

Using exact line search, $\alpha_k =$???

$p_k = 2$, $\alpha_k = 0.5$

Fibonacci Sequence

1 1 2 3 5 8 13 21 34 55 89 …

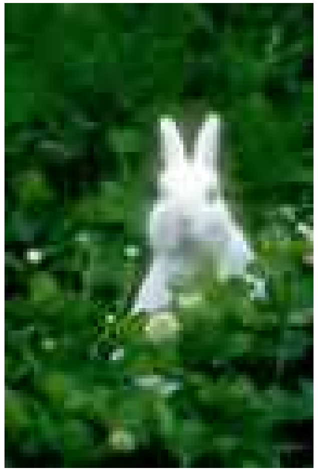


$$F_0 = F_1 = 1,\ F_n = F_{n-1} + F_{n-2}$$

for n ≥ 2

Fibonacci numbers and Nature

FLOWERS: On many plants, the number of petals is a Fibonacci number: buttercups have 5 petals; lilies and iris have 3 petals; some delphiniums have 8; corn marigolds have 13 petals; some asters have 21 whereas daisies can be found with 34, 55 or even 89 petals.



Rabbits are able to mate at the age of one month so that at the end of its second month a female can produce another pair of rabbits. Suppose that our rabbits never die and that the female always produces one new pair (one male, one female) every month from the second month on. The puzzle that Fibonacci posed was...
How many **pairs** will there be in one year?

http://www.mcs.surrey.ac.uk/Personal/R.Knott/Fibonacci/fibnat.html

Unravelling Fibonacci

Consider the sequence of the ratios of consecutive Fibonacci terms

$$\frac{1}{1},\quad \frac{2}{1},\quad \frac{3}{2},\quad \frac{5}{3},\quad \frac{8}{5},\quad \frac{13}{8},\quad \frac{21}{13},\quad \ldots$$

$$x_1 = \frac{1}{1},\quad x_2 = \frac{2}{1},\quad \ldots,\quad x_n = \frac{F(n+1)}{F(n)},\quad \ldots$$

$$x_n = \frac{F(n+1)}{F(n)}$$

$$= \frac{F(n) + F(n-1)}{F(n)}$$

$$= 1 + \frac{F(n-1)}{F(n)}$$

$$= 1 + \frac{1}{\frac{F(n)}{F(n-1)}}$$

$$\boxed{= 1 + \frac{1}{x_{n-1}}}$$

Unraveling more of Fibonacci …

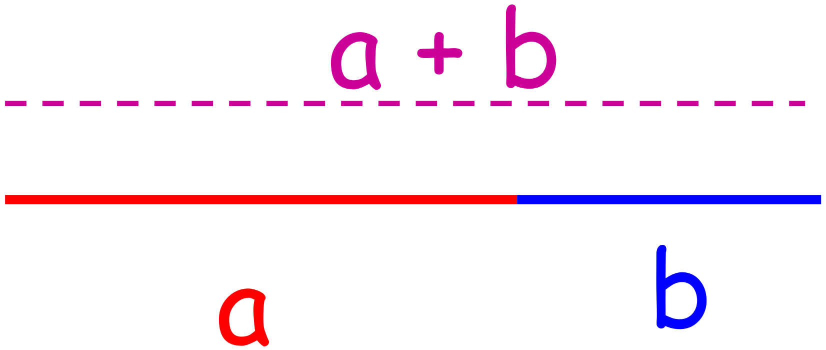
$$\lim_{n \to \infty} x_n = \lim_{n \to \infty} x_{n-1} = x$$

$$x = 1 + \frac{1}{x}$$

Solve the above equation … what do you get?

$\frac{1+\sqrt{5}}{2}$ … aka **the Golden Ratio**

The Golden Ratio



$$\frac{a}{b} = \frac{a+b}{a}$$

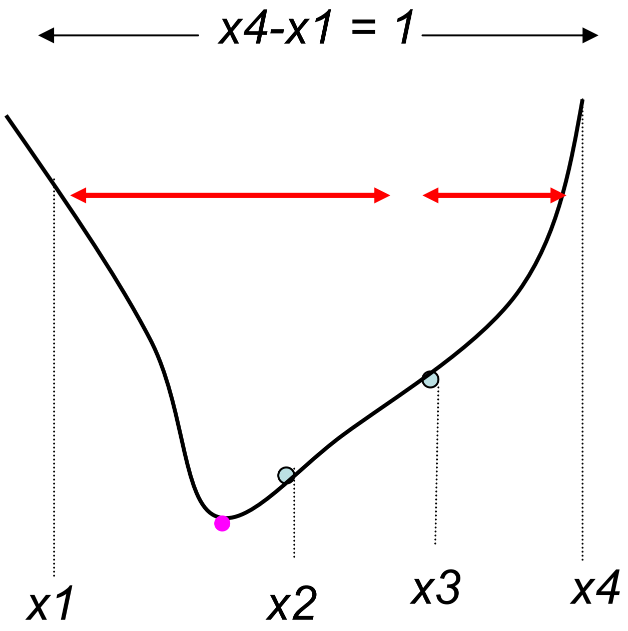
Set b=1 … what do you get??

The other solution

$$\frac{a}{b} \qquad \frac{1+\sqrt{5}}{2} = 1.618 \ldots$$

$$\frac{b}{a} \qquad \frac{\sqrt{5}-1}{2} = 0.618 \ldots$$

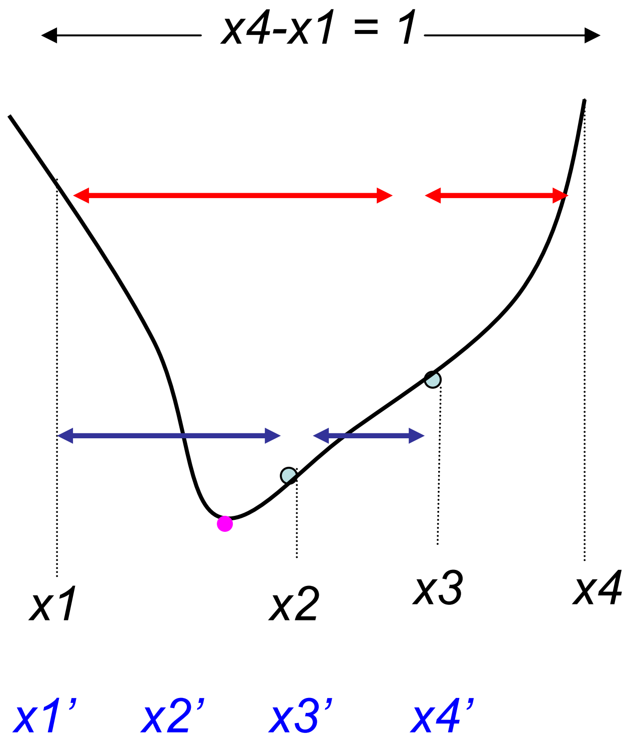
Golden Section search



Goal: Given a **convex** function, we want to shrink the interval as much as possible at each iteration, until we close in on x*.

Suppose that at some iteration in our search, we have an interval [x1, x3] that contains x*, and f(x2) < f(x3).

Golden Section search



Goal: Given a **convex** function, we want to shrink the interval as much as possible at each iteration, until we close in on x*.

Suppose that at some iteration in our search, we have an interval [x1, x3] that contains x*, and **f(x2) < f(x3)**.

Then at the next iteration, define x'1',x2',x3',x4' such that x1'=x1, x3'=x2, x4'=x3.

$$\rightarrow x3' = x1 + \tau (x3-x1) = x1 + \tau (\boxed{x4'}-x1)$$

$$= x1 + \tau \boxed{(x1 + \tau (x4-x1))} -x1)$$

$$= x1 + \tau^2 (x4-x1)$$

$$x4 - \tau (x4-x1) = x1 + \tau^2 (x4-x1)$$

$$\rightarrow (x4-x1) - \tau (x4-x1) = \tau^2 (x4-x1)$$

$$\rightarrow 0 = (\tau^2 + \tau -1)*(x4-x1)$$

Newton’s Method

```
Initial point $x_0$ and set precision to $\epsilon$
$x \rightarrow x_0$
repeat
$x \rightarrow x - [\bigtriangledown^2 f(x)]^{-1}[\bigtriangledown f(x)]$
until $| \bigtriangledown f(x)| < \epsilon$
```

In the classical Newton Method, stepsize $\alpha=1$. In other versions,

$$x \rightarrow x - \alpha_k[\bigtriangledown^2 f(x)]^{-1}[\bigtriangledown f(x)]$$

In Quasi Newton methods, an approximation is used for

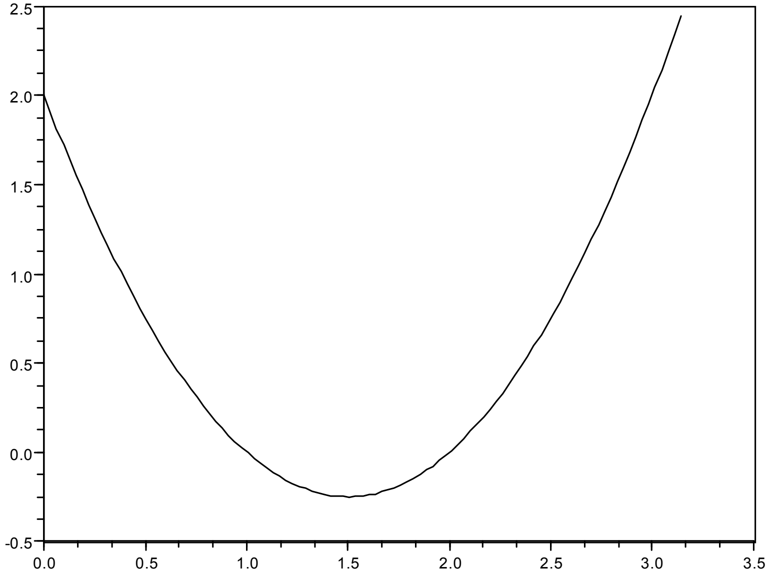
$$\bigtriangledown^2 f(x)$$

Newton's Method, a simple 1-dim Example

$f(x) = x^2 - 3x + 2$
Find Newtons method at $x_0 = 1$

$\bigtriangledown f(x) = 2x - 3,\ \bigtriangledown f(1) = -1$
$\bigtriangledown^2 f(x) = 2$

$x \rightarrow 10.5 * (-1) = 1.5$



Now try $x_0 = -2$

Newton's Method, 2-dim example

Consider the function

$$f(x_1, x_2) = \frac{1}{3}x_1^3 + \frac{1}{2}x_1^2 + 2x_1x_2 + \frac{1}{2}x_2^2 - x_2 + 9$$

$$\bigtriangledown f(x) = \begin{pmatrix} x_1^2 + x_1 + 2x_2 \\ 2x_1 + x_2 - 1 \end{pmatrix} \qquad \bigtriangledown^2 f(x) = \begin{pmatrix} 2x_1 + 1 & 2 \\ 2 & 1 \end{pmatrix}$$

$$\text{At } x = [2 \qquad -3], \quad \bigtriangledown f(x) = \begin{pmatrix} 0 \\ 5 \end{pmatrix} \qquad \bigtriangledown^2 f(x) = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$$

$$[\delta^2 f(x)]^{-1} = ???$$

Inverse of a matrix

Theorem. For any square matrix *A* of order n, we have

$$A \cdot adj(A) = \det(A) I_n.$$

For n>2,
things get more
complicated

In particular, if $\det(A) \neq 0$, then $$A^{-1} = \frac{1}{\det(A)} adj(A).$$

For a square matrix of order 2, we have $$adj\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}^T = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

which gives $$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

$$\bigtriangledown^2 f(x) = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} \rightarrow [\bigtriangledown^2 f(x)]^{-1} = \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix}$$

Steepest vs Newton

- 19th century, Cauchy
- simple
- does not require second derivatives
- low cost per iteration

❖ slower rate of convergence (linear)

❖ convergence so slow that sometimes $x_{k+1} – x_k$ is below computer precision

❖ many iterations

❖ optimal solution may never be reached

❖ 17th century, Newton

❖ expensive iterations, requires second derivatives and their inverses

❖ it can sometimes fail

- fast rate of convergence (quadratic)