Lecture 13

Finding solutions of nonlinear equations:

Secant method as a modification of Newton's method:

If we don't know the derivative, just numerically estimate it.

Algorithm: Begin w/ x_1 & x_2

Now calculate $f'(x_2) \simeq \dfrac{f(x_2) - f(x_1)}{x_2 - x_1}$

Substitute estimate into Newton method equation:

$$x_3 = x_2 - f(x_2)\,\frac{x_2 - x_1}{f(x_2) - f(x_1)}$$

convergence is fast like Newton's method, under good conditions.

Newton's method for multiple variables

Suppose we have simultaneous nonlinear equations. We can write them as

$$f_1(x_1, \dots, x_N) = 0$$

$$\vdots$$

$$f_N(x_1, \dots, x_N) = 0$$

(should be same # of functions as variables)

Suppose there is a solution @ $x_1^*, \dots, x_N^*$ so that

$$f_i(x_1^*, \dots, x_N^*) = 0 \quad \forall i \in \{1, \dots, N\}$$

then Taylor expand about solution

$$f_i(x_1^*, \dots, x_N^*) = f_i(x_1, \dots, x_N) + \sum_j (x_j^* - x_j) \frac{\partial f_i}{\partial x_j} + \dots$$

could also write in vector notation as

$$\bar{f}(\bar{x}^*) = \bar{f}(\bar{x}) + \nabla \bar{f} \cdot (\bar{x}^* - \bar{x}) + \dots$$

where $\nabla \bar{f}$ is the Jacobian matrix w/ entries $\frac{\partial f_i}{\partial x_j}$.

Since $\bar{x}^*$ was assumed to be a solution, this implies that $\bar{f}(\bar{x}^*) = 0$

$$\Rightarrow \quad 0 = \bar{f}(\bar{x}) + \nabla \bar{f} \cdot (\bar{x}^* - \bar{x})$$

$$\Leftrightarrow \quad \nabla \bar{f} \cdot (\bar{x} - \bar{x}^*) = \bar{f}(\bar{x})$$

Set $\Delta \bar{x} = \bar{x} - \bar{x}^*$

$$\Leftrightarrow \quad \nabla \bar{f} \cdot \Delta \bar{x} = \bar{f}(\bar{x})$$

this is a matrix equation which we can solve by Gaussian elimination

After doing so, we get $\Delta \bar{x}$ + then the next estimate is

$$\bar{x}' = \bar{x} - \Delta \bar{x}$$

If we cannot get Jacobian matrix, then numerically estimate, ~~&~~ we get vector secant method

Maxima + minima of functions

Finding optima is closely related to root finding.

Stick to optimization of a single function, i.e., minimum of $f(x_1, \ldots, x_N)$

functions can have more than one minimum or maximum.

distinguish between local optimum + global optimum

Standard method for computing minimum is to solve $\frac{\partial f}{\partial x_i} = 0 \quad \forall i$

if equations are linear, use Gaussian elimination

if nonlinear, use Newton's method

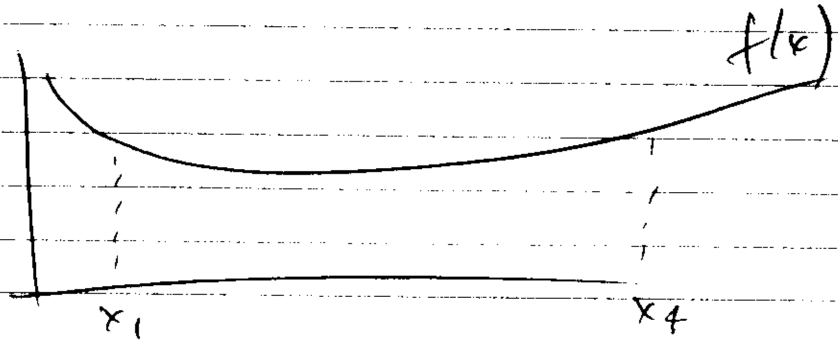
Many times we cannot calculate the derivative, so we need alternate techniques.

One method is the golden ratio search (similar to binary search)

This works only for finding the minimum of a single-variable function.

Suppose we're trying to find the minimum. This method will not distinguish between local & global minimum (just gives you some minimum)

Suppose picture is like



1st pick two points x_1 & x_4 which correspond to an interval where we would like to search for a minimum.

Now pick x_2 & x_3 such that

$$x_1 < x_2 < x_3 < x_4$$

(will give spacing later)

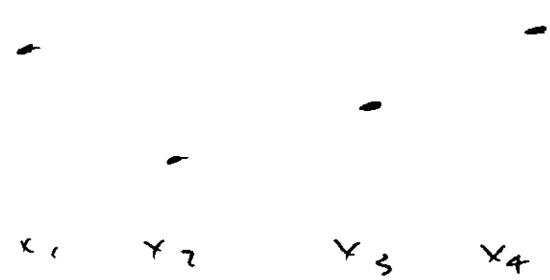
Suppose that $f(x_2) < \min\{f(x_1), f(x_4)\}$

or $f(x_3) < \min\{f(x_1), f(x_4)\}$

then we know that minimum has to be between them

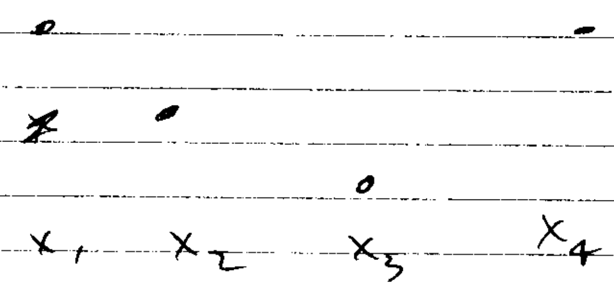
Now compare $f(x_2)$ w/ $f(x_3)$

If $f(x_2) < f(x_3)$, then picture is



& minimum should be between x_1 & x_3

Otherwise, picture B



& minimum is between x_2 & x_4

can keep repeating this process by adding another point to the three we ~~narrow~~ down to.

How should we pick x_1, x_2, x_3, x_4?

x_1 & x_4 are fixed.

Since we don't know a priori whether minimum will be on left or right side, it makes no sense to favor one over the other. So ~~the~~ x_2 & x_3 should be symmetric about midpoint

x_1 $\quad$ x_2 $\quad$ $\frac{x_1+x_4}{2}$ $\quad$ x_3 $\quad$ x_4

So pick $\frac{x_1+x_4}{2} - x_2 = x_3 - \left(\frac{x_1+x_4}{2}\right)$

$\Leftrightarrow \quad x_2 - x_1 = x_4 - x_3$

We need another equation to pin down the next location for x_2 & x_3.

For one iteration, suppose that

$x_1 \qquad x_2 \quad | \quad x_3 \qquad x_4$

If $f(x_2) < f(x_3)$, $\min\{f(x_1), f(x_4)\}$

then we would next choose x_1 & x_3

$x_1 \quad x_5 \;|\; x_2 \quad x_3$

So if we originally choose x_2 & x_3 close to midpoint, then we shave off quite a bit of space to consider but the expence occurs on the next iteration b/c the interval x [illegible] [illegible] will be larger than it should

How to make this fair?

choose the proportions of intervals on each iteration to be the same.

~~...~~ Supposing that we are always "going left", this gives the ratio ϕ_i of interval sizes to be

$$\phi_i = \frac{x_4 - x_1}{x_3 - x_1} = \frac{x_2 - x_1 + x_3 - x_1}{x_3 - x_1} = \frac{x_2 - x_1}{x_3 - x_1} + 1$$

For the next step, the ratio is

$$\phi_{i+1} = \frac{x_3 - x_1}{x_2 - x_1}$$

To make things fair for each iteration, we set the ratios equal, giving

$$\phi_{i+1} = \phi_i = \frac{1}{\phi_{i+1}} + 1$$

$$\Leftrightarrow \quad \phi_{i+1}^2 - \phi_{i+1} - 1 = 0$$

Solving this gives $\phi_{i+1} = \frac{1+\sqrt{5}}{2}$ which is the golden ratio this ϕ is known as the golden ratio..

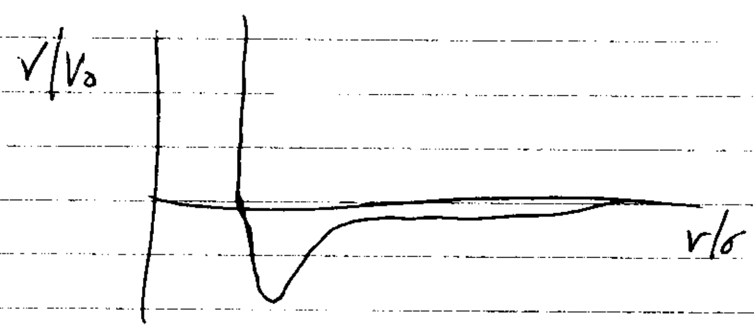
Complete algorithm

1. Pick 2 initial points x_1 & x_4.
   Calculate x_2 & x_3 according to golden ratio rule.
   Calculate $f(x)$ @ all four points
   & check that either $f(x_2) \leq \min\{f(x_1), f(x_4)\}$
   or $f(x_3) < \min\{f(x_1), f(x_4)\}$
   Choose $\varepsilon > 0$.
2. If $f(x_2) \leq f(x_3)$ then
   $x_4 = x_3$
   $x_3 = x_2$
   Pick new x_2 according to golden ratio
   & calc. $f(x_2)$
3. Else, set $x_1 = x_2$, $x_2 = x_3$,
   & get x_3 & $f(x_3)$
4. If $|x_4 - x_1| > \varepsilon$, go to 2. Else calculate $\frac{1}{2}(x_2 + x_3)$

Example of Buckingham potential:

Approximate representation of potential energy of interaction between atoms in a solid or gas as a function of distance r between them

$$V(r) = V_0 \left[\left(\frac{\sigma}{r}\right)^6 - e^{-r/\sigma} \right]$$



Two terms - short range repulsive force + longer range attractive force

No known analytic expression for "resting distance"

Bring up buckingham.py

golden ratio search cannot be generalized to functions of more than one variable.

Gauss-Newton method

To find minimum of $f(x)$,

set $f'(x) = 0$ & solve for roots

Using Newton method, we get the update rule

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

fast convergence & can be generalized to more than one variable.

If we cannot calculate 2nd derivatives then approximate by

$$x_{i+1} = x_i - \gamma f'(x_i)$$

where γ is a constant value representing a guess for $f''(x)$

Doing this is called gradient descent
- measures gradient @ x + subtracts a constant times the gradient

For $\gamma > 0$ we converge to a minimum + for $\gamma < 0$ we converge to a maximum.

Magnitude of γ controls the rate of convergence. If γ large, we go faster but could then "overshoot" the solution.

If we cannot calculate the derivative, then just numerically estimate it using two successive points

Introduction to Semidefinite Programming - a powerful analytical & numerical tool for solving optimization problems w/ semidefinite constraints - could be useful in finding the ground state of a Hamiltonian - often used in quantum information science.

Let X be a Hermitian matrix acting on a finite-dimensional complex vector space $\mathcal{X}$ & let Y be ... on $\mathcal{Y}$

Let Φ denote a linear "superoperator" which takes $X \in \mathcal{X}$ to $Y \in \mathcal{Y}$

i.e., $\Phi(X) = Y$

Φ is Hermiticity preserving if
$\Phi(X) \in \mathrm{Herm}(\mathcal{Y})$ for all $X \in \mathrm{Herm}(\mathcal{X})$

A semidefinite program is a triple (Φ, A, B):

1. Φ is Hermiticity preserving
2. $A \in \mathrm{Herm}(\mathcal{X})$ & $B \in \mathrm{Herm}(\mathcal{Y})$

Let $\langle C, D \rangle = \mathrm{Tr}\{C^{+} D\}$ be Hilbert-Schmidt inner product between matrices C & D

Define the adjoint Φ^{+} of a superoperator Φ by

$$\langle E, \Phi(F) \rangle = \langle \Phi^{+}(E), F \rangle$$

which holds for all $E \in L(\mathcal{Y})$ & $F \in L(\mathcal{X})$

Associated w/ Φ, A, B are two optimization problems, called the primal & dual

Primal	Dual
$\max \langle A, X \rangle$	$\min \langle B, Y \rangle$
subject to	subject to
$\Phi(X) = B$	$\Phi^{*}(Y) \geq A$
$X \in \mathrm{Pos}(\mathcal{X})$	$Y \in \mathrm{Herm}(\mathcal{Y})$

$\langle A, X \rangle$ is objective function

$\Phi(X) = B$ & $X \in \mathrm{Pos}(\mathcal{X})$

are constraints

$C \leq D$ means $D - C$ is PSD

$\mathrm{Pos}(\mathcal{X})$ is the set of positive semidefinite operators acting on $\mathcal{X}$

Primal & dual problems have a special relationship

An operator $X \in \mathrm{Pos}(\mathcal{X})$ for which $\Phi(X) = B$ is primal feasible

Let $\mathcal{A} = \{X \in \mathrm{Pos}(\mathcal{X}) : \Phi(X) = B\}$

$Y \in \mathrm{Herm}(\mathcal{Y})$ satisfying

$\Phi^{+}(Y) \geq A$ is dual feasible

Let $\mathcal{B} = \{ Y \in \mathrm{Herm}(\mathcal{Y}) : \Phi^{+}(Y) \geq A \}$

primal optimum is $\alpha = \sup_{X \in \mathcal{A}} \langle A, X \rangle$

dual optimum is $\beta = \inf_{Y \in \mathcal{B}} \langle B, Y \rangle$

α + β might be finite or infinite

$\alpha = -\infty$ if $\mathcal{A} = \phi$

$\beta = \infty$ if $\mathcal{B} = \phi$

If $X \in \mathcal{A}$ satisfies $\langle A, X \rangle = \alpha$

then X is ~~primal optimal~~

optimal primal solution

If $Y \in \mathcal{B}$ satisfies $\langle B, Y \rangle = \beta$

then Y is optimal dual solution

Example:

Take $\Phi(X) = \mathrm{Tr}\{X\}$

$B = 1$

Then

Primal	Dual
max $\langle A, X\rangle$	minimize y
subject to $\mathrm{Tr}\{X\} = 1$	subject to $yI \geq A$
$X \in \mathrm{Pos}(\mathcal{X})$	$y \in \mathbb{R}$

adjoint of Trace is multiplication by identity

---

semidefinite programming has a powerful duality theory.

Begin w/ weak duality:

For every semidefinite program (Φ, A, B) it holds that $\alpha \leq \beta$.

Proof: trivial if $\mathcal{A} = \phi$ or $\mathcal{B} = \phi$

Then For all $X \in \mathcal{A}$ and $Y \in \mathcal{B}$ we have

$$\langle A, X\rangle \leq \langle \Phi^{+}(Y), X\rangle = \langle Y, \Phi(X)\rangle$$

$$= \langle Y, B\rangle \Rightarrow \alpha \leq \beta$$