we shall require that:

1) the algorithm makes progress to the sol'n at every step

2) The steps are never too large.

Thus we require:

$$\|\underset{\sim}{f}(\underset{\sim}{x}^{(k+1)})\|_2 < \|\underset{\sim}{f}(\underset{\sim}{x}^{(k)})\|_2$$

and that

$$\|\underset{\sim}{x}^{(k+1)} - \underset{\sim}{x}^{(k)}\|_2 < \delta$$

where δ is picked by the algorithm

if the truncated Taylor series (including the linear term) is a <u>good</u> approx to the function, we make δ large, if not we reduce δ

Suppose we have :

$$\underset{\sim}{x}^{(k+1)} = \underset{\sim}{x}^{(k)} + \underset{\sim}{p}$$

correction at step k

From Newton's method we expect :

$$\underset{\sim}{p} = -\left(\underset{\approx}{J}^{(k)}\right)^{-1} \underset{\sim}{f}^{(k)}$$

but this may not satisfy $\|\underset{\sim}{p}\|_2 < \delta$

In addition, if $\underset{\approx}{J}^{(k)}$ is singular, $\underset{\sim}{p}$ may not exist!

Instead, let's solve the constrained optimization problem :

$$\min_{\underset{\sim}{p}} \| \underset{\approx}{J}^{(k)} \underset{\sim}{p} + \underset{\sim}{f}^{(k)} \|_2$$

subject to $\|\underset{\sim}{p}\|_2 < \delta$

This works even for singular $\underset{\approx}{J}^{(k)}$!

This is a bit different from the linear regression problem. Linear regression is an <u>unconstrained</u> optimization problem, what we have is a <u>constrained</u> problem.

We'll look into solving this in a bit. If we have such a solver, we would

1) if $\underset{\sim}{f}(\underset{\sim}{x}^{k}) \approx 0$ then stop

2) calc. $\underset{\sim}{p}$ via constrained optimization solution

3) If $\| \underset{\sim}{f}(x^{(k)} + \underset{\sim}{p}) \|_2 < \| f(\underset{\sim}{x}^{(k)}) \|_2$ then accept step & go to 1. You may also want to increase δ.

4) If not, then reduce δ & go back to step 2.

This sort of algorithm will reliably find roots, and will be trapped by local minima.

Key: start close to the answer!

Optimization

In optimization we are trying to minimize or maximize an objective function, usually subject to a set of constraints

This is known as a constrained optimization problem

If there are no constraints, then we have an unconstrained optimization problem.

Example: Suppose we are designing soup cans. We want to minimize the metal usage for volume enclosed, e.g., minimize surface/volume ratio

Obviously, a sphere would do this, but it's difficult to make & open.

Instead, we choose a cylindrical shape:

Metal Area $= \underbrace{2\pi r^2}_{\text{2 ends}} + \underbrace{2\pi r h}_{\text{sides}} = A$

Volume $= \pi r^2 h = V$

We wish to minimize A subject to the constraint that $V = 16$ oz (say)

This is a 2-D non-linear constrained optimization problem

We can convert this to a 1-D unconstrained problem by recognizing:

$$h = \frac{V}{\pi r^2}$$

$$\therefore \quad A = 2\pi r^2 + \frac{2V}{r}$$

Which has the sol'n $r = 1.08$, $h = 2.17$

$$A = 22.14$$

Actually, cans don't fit this ideal, as there are additional constraints such as labelling and packing considerations.

We have the general problem:

minimize $F(\underset{\sim}{x}) \equiv F(x_1, x_2 \ldots x_n)$

over some domain S in n-dim. space

For the can problem,

$$F(x_1, x_2) = 2\pi x_1^2 + 2\pi x_1 x_2$$

subject to $\underset{\sim}{x} \in S = \{(x_1, x_2) \mid 2\pi x_1^2 x_2 = V\}$

If $S \equiv \mathbb{R}_N$ (all space) then problem is unconstrained

$\underset{\sim}{x} \in S$ are the feasible solutions

Least squares data fitting is an unconstrained optimization problem.

If we have some point $\underset{\sim}{x}$ s.t. $\underset{\sim}{\nabla} F = 0$

then $\underset{\sim}{x}$ is a critical pt

It may be a minimum, a maximum, or a saddle point!

145

We define a point $\underset{\sim}{x}^*$ to be a <u>local minimum</u> if:

$$\underset{\sim}{x}^* \varepsilon S \quad \text{s.t.} \quad F(\underset{\sim}{x}^*) < F(\underset{\sim}{x}^* + \underset{\sim}{\delta})$$

$\underset{\sim}{x}^*$ is a <u>global</u> <u>minimum</u> if

$$F(\underset{\sim}{x}^*) < F(\underset{\sim}{x}) \; ; \; \underset{\sim}{x} \; \varepsilon \; S \qquad (\text{all } \underset{\sim}{x})$$

You can't guarantee to find the global minimum! (OR - interval techiques <u>do</u> allow this over some domain)

One-Dim. problems

There are three basic approaches analogous to the three root finding techniques. They are:

1) Newton's Method

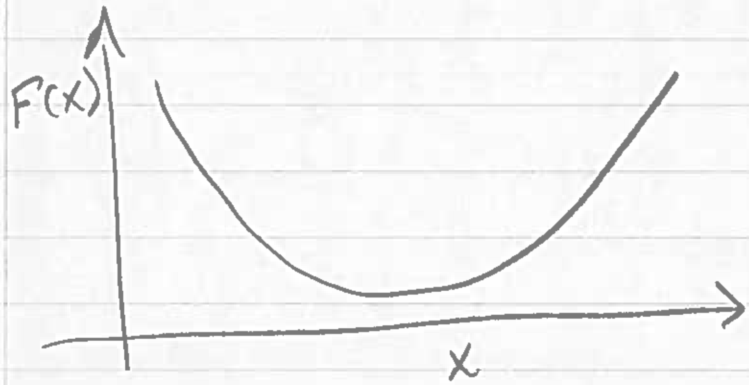
2) Successive Parabolic Interpolation

3) Golden Search (Fibonacci Search)

146

We start with Newton's method

We have the one-D function $F(x)$:



It is natural to approximate this with a parabola. We thus truncate its Taylor series after the quadratic term, and use the approx. to get the next estimate of the critical point:

$$F(x) = F(x_k) + (x - x_k)F'(x_k) + \frac{1}{2}(x - x_k)^2 F''(x_k) + \ldots$$

We seek x^* where $F'(x^*) = 0$

Thus:

$$x_{k+1} = x_k - \frac{F'(x_k)}{F''(x_k)}$$

This is virtually identical to root solving via Newton's method ⇒ in fact it ~~is~~ root solving, just the root of $F'(x)$

Successive Parabolic Interpolation is very similar to the secant method for root solving.

Instead of using two points to get a line, we use three points to get a parabola!

Thus if we have x_k, x_{k-1}, x_{k-2}

we fit $y = ax^2 + bx + c$ to the points:

$$(x_k, F(x_k)), (x_{k-1}, F(x_{k-1})), (x_{k-2}, F(x_{k-2}))$$

Once we have a, b & c we seek the critical pt.

$$\frac{dy}{dx} = 0 = 2ax + b$$

Thus:

$$x_{k+1} = \frac{-b}{2a}$$

and we iterate forward keeping the last three points

The rate of convergence is 1.324

This is superlinear, but not as fast as Newton's method ($r=2$)

Both techniques go to critical pts - either maxima or minima - and have convergence problems if you don't start close enough to the correct answer!

The Golden Search (modified Fibonacci search) is analogous to root finding using bisection, In this case rather than knowing some interval $[a,b]$ where $f(a)f(b)<0$ as in bisection, here we require that the function $F(x)$ be unimodal over $[a,b]$

This means that (if we are looking for a minimum):

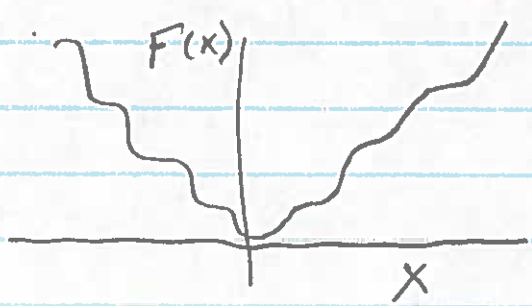
$$F'(x) \begin{cases} < 0 & x < x^* \\ > 0 & x > x^* \end{cases}$$

Note that F'' may change sign in this interval! If $F''=0$ Newton's method (and successive parabolic interpolation) will fail, but the golden search is unaffected!

An example of such a function:

$$F(x) = x^2 + a\cos(\omega x)^2$$

provided $\omega^2 a < 1$ there is only one critical point at $x^* = 0$

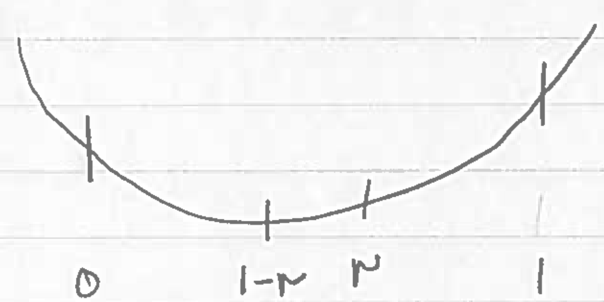


For large x, however, F'' is both positive and negative. Unless Newton's method is started close to the min. it will not converge!

The golden search relies on evaluating the function at points within the interval $[a, b]$ and determining in which part of the sub-interval the minimum lies.

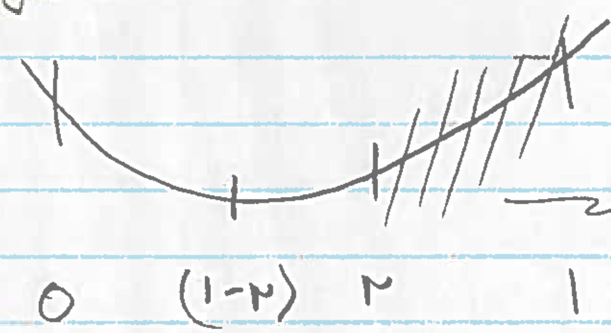
Suppose we have the interval $[0, 1]$ over which the function is unimodal, and which contains the minimum.

We evaluate the function at two additional points in the interior $(1-r), r$ s.t. $r > 0.5$. Thus:



Now if $F(r) < F(1-r)$ then the minimum will lie in the interval: $[(1-r), 1]$ and if $F(r) > F(1-r)$ then it will lie in the interval $[0, r]$.

We discard the balance of the original interval and do it again:



$F(r) > F(1-r)$ so right part is discarded!

The trick is to choose r so that in the next step we only need one new evaluation rather than two!

This means that we want the point $(1-r)$ on the left side of the original interval to map onto the point r on the right hand side of the new subinterval!

Thus we require:

$$\underset{\substack{\text{left hand}\\\text{pt original}\\\text{interval}}}{(1-r)} = \underset{\substack{\text{right hand}\\\text{point of}\\\text{sub interval}}}{(r)}\ \underset{\substack{\text{width of sub-}\\\text{interval}}}{(r)}$$

Thus $r^2 + r - 1 = 0$

$$r = \frac{\sqrt{5}-1}{2} = 0.6180$$

Golden Ratio

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

You get the same value if you discard the left side of the interval and map the old rhs point onto the new left hand side point.

each iteration discards 38% of the current interval, thus:

$$\frac{|e_{i+1}|}{|e_i|} = 0.618$$

So we have <u>linear</u> rate of convergence, but C is higher than in bisection.

Run Matlab Example!