7. SOLVING NONLINEAR EQUATIONS

- Nonlinear equations: Bisection, Fixed-point, and Newton methods
- Nonlinear equations: Multivariate Newton
- Nonlinear least squares: Newton, Gauss-Newton, and Levenberg-Marquardt

- Univariate polynomial equations: Laguerre's method
- Univariate polynomial equations: Sturm sequences

Class web site:

`http://www.di.ens.fr/~brette/calculscientifique/index.htm`

`http://www.di.ens.fr/~ponce/scicomp/notes.pdf`

Bisection

- Suppose that we know a and $b \in \mathbb{R}$ such that $f(a)f(b) < 0$. [BRACKETING]
- Can we find a zero of f?

- By continuity, f must be zero in some point $x_0 \in]a_1 b[$

  $\Longrightarrow$ Compute $f(c)$ where $c = \frac{a+b}{2}$

  $\Longrightarrow$ If $f(c) \begin{array}{l} > 0 \text{ there must be a zero on }]cb[\\ < 0 \text{ there must be a zero on }]ab[\end{array}$

  etc...

Recursive procedure that subdivides the half interval containing a root: bisection.

- When a and b are given, bisection always converges to a root. The speed of concergence is linear: $|x_k - x_0| \leq \frac{1}{2^k}|b - a|$.

- *Problem:* How do you find a bracketing interval?

```
// bisection method
function [zero,res,niter]=bisection(fun,a,b,tol,nmax)
x = [a, (a+b)*0.5, b];
fx = fun(x);
if fx(1)*fx(3)>0
    error(' The sign of FUN at the extrema of the interval must be diff
    elseif fx(1) == 0, zero = a; res = 0; niter = 0; return
    elseif fx(3) == 0, zero = b; res = 0; niter = 0; return
end
niter = 0;
I = (b - a)*0.5;
while I >= tol & niter <= nmax
   disp(sprintf('i=%d x=%20.15f', niter,x))
   niter = niter + 1;
   if sign(fx(1))*sign(fx(2)) < 0
      x(3) = x(2);
      x(2) = x(1)+(x(3)-x(1))*0.5;
      fx = fun(x);
      I = (x(3)-x(1))*0.5;
   elseif sign(fx(2))*sign(fx(3)) < 0
      x(1) = x(2);
      x(2) = x(1)+(x(3)-x(1))*0.5;
      fx = fun(x);
      I = (x(3)-x(1))*0.5;
   else
       x(2) = x(find(fx==0));
       I = 0;
   end
end
if niter > nmax
    fprintf(['bisection stopped without converging to the desired toler
    'because the maximum number of iterations was reached\n']);
end

zero = x(2);
x = x(2); res = fun(x);
endfunction
```

Bisection Test

Find a root of $f(x) = \tan(\dfrac{x}{4}) - 1$ in the interval [2,4]

$$\frac{x}{4} = \frac{\pi}{4} \Leftrightarrow x = \pi$$

One new digit of accuracy every few iterations.

Fixed-Point Method

Given a continuous function $f : \mathbb{R} \to \mathbb{R}$, what is a *fixed point* x^* of f such that $x^* = f(x^*)$?

This is a point where the curve $y = f(x)$ intersects the line $y = x$.

A plausible method for finding such a point is to use *fixed-point iterations*

$$x_{i+1} = f(x_i).$$

When does it converge?

Fixed-Point Method (2)

• Assume f is C^1, and $|f'(x^*)| < 1$. Then, by the mean value theorem,

$$e_{i+1} = x_{i+1} - x^* = f(x_i) - f(x^*) = f'(\theta_i)(x_i - x^*) = f'(\theta_i)e_i$$

for some θ_i in the interval between x_i and x^*.

• In particular, if $|f'(x^*)| < 1$, there exists a constant C and some interval centered I in x^* such that, if $x_0 \in I$, we have $|f'(\theta_i)| \leq C < 1$ for all $i \geq 0$, and thus

$$|e_{i+1}| \leq C|e_i| \leq \ldots \leq C^k|e_0|.$$

• In this case, the fixed-point method is guaranteed to converge with linear convergence rate.

• Suppose further that $|f'(x^*)| = 0$. Then

$$e_{i+1} = f(x_i) - f(x^*) = \frac{1}{2}f''(\eta_i)(x_i - x^*)^2 = \frac{1}{2}f''(\eta_i)e_i$$

for some η_i in the interval between x_i and x^*. Thus,

$$lim_{k\to+\infty}\frac{e_{i+1}}{e_i} = \frac{1}{2}f''(x^*).$$

The method has *quadratic* convergence in this case.

Newton's Method

- Suppose you are at a point x_i on the curve.

- How can you move to a close-by root x_0?

- *Idea:* Pretend the curve and its tangent in x_i coincide and iterate...

$$0 = f(x_{i+1}) = f(x_i) + (x_{i+1} - x_i)f'(x_i) + \varepsilon(x_{i+1} - x_i)^2$$

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

Newton's Method (2)

What is the rate of convergence? Define $\varepsilon_i = x_i - x_0$.

By definition

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

But

$$\begin{cases} f(x_i) = f(x_0) + \varepsilon_i f'(x_0) + \frac{1}{2}\varepsilon_i^2 f''(x_0) + ... \\ f'(x_i) = f'(x_0) + \varepsilon_i f''(x_0) + ... \end{cases}$$

Substituting in (1) yields

$$\begin{aligned} \varepsilon_{i+1} &= \frac{1}{f'(x_i)}\left[\varepsilon_i f'(x_i) - \varepsilon_i f'(x_0) - \frac{1}{2}\varepsilon_i^2 f''(x_0) + ...\right] \\ &= \frac{1}{f'(x_0) + ...}\left[\varepsilon_i f'(x_0) + \varepsilon_i^2 f''(x_0) - \varepsilon_i f'(x_0) - \frac{1}{2}\varepsilon_i^2 f''(x_0) + ...\right] \\ &\Downarrow \\ \varepsilon_{i+1} &\approx \frac{1}{2}\varepsilon_i^2 \frac{f''}{f'}(x_0) \end{aligned}$$

$\Downarrow$

Quadratic convergence!

Newton's Method (3)

- In general, the error is not guaranteed to decrease at every iteration.

- In fact, the method can die because of local extrema

$$\frac{f}{f'}(x_{i+1}) \to \infty$$

- Theorem: When f satisfies a bunch of difficult-to-verify conditions, then

$$|x_{i+1} - x_0| \leq k|x_i - x_0|^2$$

$\Downarrow$

- In other words, under these assumptions
  - Newton will converge to a root
  - With a quadratic rate of convergence

Newton's Method (4): When do you stop?

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

- When $|x_{i+1} - x_i|$ is small enough

  - Consider $\begin{cases} f(x) = \tan x \\ f'(x) = \dfrac{1}{\cos^2 x} \end{cases}$

  - $\dfrac{f}{f'}(x) = \sin x \cos x \approx -10^{-6}$

  $\Downarrow$

$$\begin{cases} \text{zeros of } f : x = 0, \pi, 2_\pi ... \\ \text{zeros of } \dfrac{f}{f'} : x = 0, \dfrac{\pi}{2}, \pi, \dfrac{3\pi}{2} ... \end{cases}$$

- When $|f(x_i)|$ is small enough

  - Consider $f(x) = (x-1)^{100}$

  - $f(1.1) = (0.1)^{100} = 10^{-100}$

  $\Downarrow$

  but the actual root is in 1.

Combining Bisection and Newton's Method

<u>Idea:</u>

- Start with a bracketing interval.
- Try Newton.
- If you get out of the bracketing interval, then use bisection instead.
- Iterate.

When do you stop?

- When the current bracketing interval is small enough,

or

- When $|f(x_i)|$ is small enough,

or

- When you have done too much work (too many iterations).

```
//   Newton plus bisection
//   Iteration terminates as soon as x is with tolx of a true zero or
//   if |f(x)|<= tolf or after nEvalMax f-evaluations
function [x,fx,nEvals,aF,bF]=GlobalNewton(fName,fpName,a,b,tolx,tolf,nE
fa  = fName(a);
fb  = fName(b);
if fa*fb>0, disp('Initial interval not bracketing.'); return end
x   = a; fx  = fName(x); fpx = fpName(x);
disp(sprintf('%20.15f  %20.15f    %20.15f',a,x,b))
nEvals = 1;
while (abs(a-b)>tolx) & (abs(fx) > tolf) & ((nEvals<nEvalsMax) | (nEval
      //[a,b] brackets a root and x = a or x = b.
      if StepIsIn(x,fx,fpx,a,b)
      //Take Newton Step
  disp('Newton')
      x   = x-fx/fpx;
  else
      //Take a Bisection Step:
  disp('Bisection')
  x = (a+b)/2;
  end
      fx  = fName(x);
      fpx = fpName(x);
  nEvals = nEvals+1;
      if fa*fx<=0
 // There is a root in [a,x]. Bring in right endpoint.
         b  = x;
         fb = fx;
      else
// There is a root in [x,b]. Bring in left endpoint.
        a  = x;
fa = fx;
      end
      disp(sprintf('%20.15f  %20.15f    %20.15f',a,x,b))
   end
   aF = a;
   bF = b;
endfunction
```

Multivariate Root Finding

Problem: Given the functions $f_i : \mathbb{R}^n \to R(i = 1, ..., n)$, what are the values of $x_1 ... x_n$

such that

$$\begin{cases} f_1(x_1, x_2, ..., x_n) = 0 \\ f_2(x_1, x_2, ..., x_n) = 0 \\ \dots \\ f_n(x_1, x_2, ..., x_n) = 0 \end{cases}$$

Ex:

$$\begin{cases} a_1 x + b_1 y + c_1 = 0 \\ a_2 x + b_2 y + c_2 = 0 \end{cases}$$

Ex:

$$\begin{cases} (\frac{x - x_1}{a_1})^2 + (\frac{y - y_1}{b_1})^2 - 1 = 0 \\ (\frac{x - x_2}{a_2})^2 + (\frac{y - y_2}{b_2})^2 - 1 = 0 \end{cases}$$

In general, there is a finite number of solutions.

Newton's Method: Multivariate Case

What are the solutions of $f_1(\vec{x}) = \ldots = f_p(\vec{x}) = 0$?

- Taylor expansion of f_i in the neighborhood of x:

$$\begin{aligned} f_i(\vec{x} + \delta\vec{x}) &= f_i(\vec{x}) + \delta x_1 \frac{\partial f_i}{\partial x_1}(\vec{x}) + \ldots + \delta x_q \frac{\partial f_i}{\partial x_q}(\vec{x}) + O(|\delta\vec{x}|^2) \\ &\approx f_i(\vec{x}) + \nabla f_i(\vec{x}) \cdot \delta\vec{x}, \end{aligned}$$

where $\nabla f_i(\vec{x}) = (\partial f_i/\partial x_1, \ldots, \partial f_i/\partial x_q)^T$ is the *gradient* of f_i at the point $\vec{x}$

- This can be rewritten as

$$\vec{f}(\vec{x} + \delta\vec{x}) \approx \vec{f}(\vec{x}) + \mathcal{J}_{\vec{f}}(\vec{x})\delta\vec{x} = \vec{0},$$

where $\mathcal{J}_{\vec{f}}(\vec{x})$ is the *Jacobian* of $\vec{f} = (f_1, \ldots, f_n)^T$—that is, the $p \times q$ matrix

$$\mathcal{J}_{\vec{f}}(\vec{x}) \stackrel{\text{def}}{=} \begin{pmatrix} \nabla f_1^T(\vec{x}) \\ \ldots \\ \nabla f_p^T(\vec{x}) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\vec{x}) & \ldots & \frac{\partial f_1}{\partial x_q}(\vec{x}) \\ \ldots & \ldots & \ldots \\ \frac{\partial f_p}{\partial x_1}(\vec{x}) & \ldots & \frac{\partial f_p}{\partial x_q}(\vec{x}) \end{pmatrix}.$$

- **When** $p = q$: Iterate

$$\vec{x} = \vec{x} - \mathcal{J}_{\vec{f}}^{-1}(\vec{x})\, \vec{f}(\vec{x}).$$

- Quadratic convergence rate when it converges.

Newton's Method for Nonlinear Least Squares

When $p > q$, define

$$E(\vec{x}) \stackrel{\text{def}}{=} \frac{1}{2}|\vec{f}(\vec{x})|^2 = \frac{1}{2}\sum_{i=1}^{p} f_i^2(\vec{x}),$$

and use Newton's method to find a local minimum of E as a zero of its gradient $\vec{F}(\vec{x}) = \nabla E(\vec{x})$.

- A simple calculation shows that

$$\vec{F}(\vec{x}) = \mathcal{J}_{\vec{f}}^T(\vec{x})\vec{f}(\vec{x}),$$

and differentiating this expression shows in turn that the Jacobian of $\vec{F}$ is

$$\mathcal{J}_{\vec{F}}(\vec{x}) = \mathcal{J}_{\vec{f}}^T(\vec{x})\mathcal{J}_{\vec{f}}(\vec{x}) + \sum_{i=1}^{p} f_i(\vec{x})\mathcal{H}_{f_i}(\vec{x}),$$

where $\mathcal{H}_{f_i}(\vec{x})$ denotes the *Hessian* of f_i—that is, the $q \times q$ matrix of second derivatives

$$\mathcal{H}_{f_i}(\vec{x}) \stackrel{\text{def}}{=} \begin{pmatrix} \frac{\partial^2 f_i}{\partial x_1^2}(\vec{x}) & \ldots & \frac{\partial^2 f_i}{\partial x_1 x_q}(\vec{x}) \\ \ldots & \ldots & \ldots \\ \frac{\partial^2 f_i}{\partial x_1 x_q}(\vec{x}) & \ldots & \frac{\partial^2 f_i}{\partial x_q^2}(\vec{x}) \end{pmatrix}.$$

- The term $\delta\vec{x}$ in Newton's method satisfies $\mathcal{J}_{\vec{F}}(\vec{x})\delta\vec{x} = -\vec{F}(\vec{x})$. Equivalently, $\delta\vec{x}$ is the solution of

$$[\mathcal{J}_{\vec{f}}^T(\vec{x})\mathcal{J}_{\vec{f}}(\vec{x}) + \sum_{i=1}^{p} f_i(\vec{x})\mathcal{H}_{f_i}(\vec{x})]\delta\vec{x} = -\mathcal{J}_{\vec{f}}^T(\vec{x})\vec{f}(\vec{x}).$$

Variants of Newton's Method for Nonlinear Least Squares

Gauss-Newton. What is the value of $\delta\vec{x}$ that minimizes $E(\vec{x}+\delta\vec{x})$ for a given value of $\vec{x}$?

$$E(\vec{x}+\delta\vec{x}) = |\vec{f}(\vec{x}+\delta\vec{x})|^2 \approx |\vec{f}(\vec{x}) + \mathcal{J}_{\vec{f}}(\vec{x})\delta\vec{x}|^2.$$

• The adjustment $\delta\vec{x}$ can be computed as the solution of $\mathcal{J}_{\vec{f}}^{\dagger}(\vec{x})\delta\vec{x} = -\vec{f}(\vec{x})$ or, equivalently, according to the definition of the pseudoinverse,

$$\mathcal{J}_{\vec{f}}^{T}(\vec{x})\mathcal{J}_{\vec{f}}(\vec{x})\delta\vec{x} = -\mathcal{J}_{\vec{f}}^{T}(\vec{x})\vec{f}(\vec{x}).$$

• This is Newtown where the Hessians $\mathcal{H}_{f_i}$ are taken equal to zero. This is justified when the residuals are small. Nearly quadratic convergence close to a solution.

Levenberg-Marquardt.

• Take the increment $\delta\vec{x}$ to be the solution of

$$[\mathcal{J}_{\vec{f}}^{T}(\vec{x})\mathcal{J}_{\vec{f}}(\vec{x}) + \mu\mathrm{Id}]\delta\vec{x} = -\mathcal{J}_{\vec{f}}^{T}(\vec{x})\vec{f}(\vec{x}),$$

where the parameter μ is allowed to vary at each iteration.
• This is Newton where the term involving the Hessians is this time approximated by a multiple of the identity matrix. The Levenberg–Marqardt algorithm has convergence properties comparable to its Gauss–Newton cousin, but it is more robust.

Real Algebraic Problems

Example 1: What are the real roots of $x^4 - 3x^3 + x^2 - 5x + 1$?

Example 2: Do the surfaces defined in $\mathbb{R}^3$ by $x^2 - 5xy + 3z^3 = 0$ and $-2x^3 + y^3 + 2xyz - 1 = 0$ intersect?

Example 3: If the intersection is not empty, how does it look?

Example 4: When is the ellipse defined by

$$\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} - 1 = 0$$

inside the unit circle centered at the origin?

Examples 1 and 3 are *display* problems, and example 2 is a *query* problem, and example 4 is a *constraint* problem.

Query and constraint problems can be reduced to *quantifier elimination* problems *over the reals*:

- Ex. 2: $(\exists x)(\exists y)(\exists z)[x^2 - 5xy + 3z^3 = 0 \text{ and } -2x^3 + y^3 + 2xyz - 1 = 0]$.
- Ex. 4: $(\forall x)(\forall y)[b^2(x-x_0)^2 + a^2(y-y_0)^2 - a^2b^2 = 0 \Longrightarrow x^2 + y^2 - 1 \leq 0]$.

Display problems may reduce to root finding but may also boil down to determining the topology of a set of varieties. Quantifier elimination plays a role there too.