We can treat this iteratively, starting at x_0, and finding $x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$. This leads to the algorithm:

- **Initialise:** Choose x_0 as an initial guess.
- **Iterate** until the absolute difference $|x_i - x_{i-1}| \approx 0$
  - Set $x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$.

Example: $f(x) = x^2 - 2$.
Solution: $f'(x) = 2x$ so

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

See slide for example starting at $x_0 = 0.5$. □

Compare with bisection method: Start at $a = 1/2$ and $b = 2$ to get to $a = 1.34375$ and $b = 1.4375$ at the fifth step. This produces an absolute error of 0.02688 or 1.9%. The absolute error in the Newton case is 0.00020 which is 2 orders of magnitude smaller.

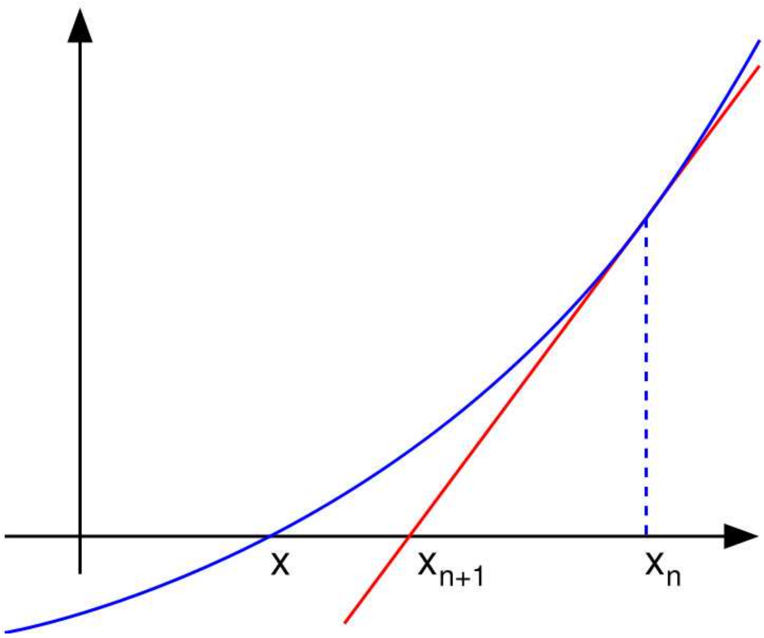

Figure 2: The idea behind the Newton's method. Starting at x_n, we find the tangent line (red) and calculate the point it intercepts the x-axis. This point of intercept is x_{n+1}.

There are, of course, many other root finding methods. We list a few of them here (not examinable).

Secant method (`http://en.wikipedia.org/wiki/Secant_method`):

- Newton's method with a finite difference instead of the derivative
- Neither computation, nor existence of a derivative is required
- However, the convergence is slower (approximately, $\alpha = 1.6$)

False position method (`http://en.wikipedia.org/wiki/False_position_method`):

- Always retains one point on either side of the root
- Faster than the bisection and more robust than the secant method

Muller's method (`http://en.wikipedia.org/wiki/Muller's_method`):

- Quadratic (instead of linear) interpolations
- Faster convergence than with the secant method
- Roots may be complex (in addition to reals)

5 Numerical linear algebra

Numerical linear algebra is one of the cornerstones of modern mathematical modelling. Topics as important as solving systems of ordinary differential equations (arising in engineering, economics, physics, biotech, etc), to network analysis (telecoms, sociologic, epidemiology), internet search, data mining and many more rely on linear algebra.

These days, applied linear algebra and numerical linear algebra are virtually interchangeable — problems of all sizes are routinely solved numerically and rely on a wealth of mathematical and computational insight.

We'll start out with a brief review of topics that you should be somewhat familiar with.

5.1 Review

Let $\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ $\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$ be vectors. The *inner* or *dot product* of $\mathbf{a}$ and $\mathbf{b}$ is the scalar $c = \mathbf{a} \bullet \mathbf{b} \equiv \mathbf{a}^\mathsf{T}\mathbf{b} = \sum_{i=1}^{n} a_i b_i$. The dot product is also called *multiplication* of vectors.

The *norm* or *magnitude* of a vector $\mathbf{a}$ is $||\mathbf{a}|| = \sqrt{a \cdot a} = \sqrt{\mathbf{a}^\mathsf{T}\mathbf{a}} = \sqrt{a_1^2 + \ldots a_n^2}$.

The *product* of an $m \times n$ matrix $\mathbf{A} = \begin{bmatrix} A_{11} & \ldots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \ldots & A_{mn} \end{bmatrix}$ and an $n \times 1$ (n-dimensional) vector $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ is the m-dimensional vector $\mathbf{y} = \mathbf{A}\mathbf{x}$ with the elements $y_i = \sum_{j=1}^{m} A_{ij} x_j$.

The *product* of a $k \times m$ matrix $\mathbf{A}$ and an $m \times n$ matrix $\mathbf{B}$ is the $k \times n$ matrix $\mathbf{C} = \mathbf{AB}$ with the elements $C_{ij} = \sum_{\alpha=1}^{m} A_{i,\alpha} B_{\alpha,j}$

The *outer product* of an m-dimensional vector $\mathbf{a}$ with an n-dimensional vector $\mathbf{b}$ is the $m \times n$ matrix

$$\mathbf{a}\mathbf{b}^\mathsf{T} \equiv \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \begin{bmatrix} b_1 & b_2 & \ldots & b_n \end{bmatrix} = \begin{bmatrix} a_1 b_1 & a_1 b_2 & \ldots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \ldots & a_2 b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_m b_1 & a_m b_2 & \ldots & a_m b_n \end{bmatrix}$$

The *identity matrix* of size n, $\mathbf{I}_n$, is the $n \times n$ matrix with (i,j)th entry $= 0$ if $i \neq j$ and 1 if $i = j$.

The *inverse* of a square matrix $\mathbf{A}$ of size n is the square matrix $\mathbf{A}^{-1}$ such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n = \mathbf{A}^{-1}\mathbf{A}$. When such a matrix exists, A is called *invertible* or *non-singular.* $\mathbf{A}$ is *singular* if no inverse exists. Finding the inverse of $\mathbf{A}$ is typically difficult.

The *determinant* of an $n \times n$ matrix $\mathbf{A}$, written $det(\mathbf{A}) = \begin{vmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \dots & A_{nn} \end{vmatrix}$, is given by a somewhat complex formula that we need not reproduce here (look it up at `http://en.wikipedia.org/wiki/Determinant`). For $n = 2$, $det(A) = A_{11}A_{22} - A_{21}A_{12}$. For $n = 3$, $det(A) = A_{11}A_{22}A_{33} - A_{31}A_{22}A_{13} + A_{12}A_{23}A_{31} - A_{32}A_{23}A_{11} + A_{13}A_{21}A_{32} - A_{33}A_{21}A_{12}$.

Example: Find the determinant of $\mathbf{A} = \begin{pmatrix} 3 & 5 \\ 1 & -1 \end{pmatrix}$.

Solution: From above, $det(\mathbf{A}) = |\mathbf{A}| = 3.{-1} - 1.5 = -3 - 5 = -8$. □

It is worth recalling a few properties of the determinant (as listed on the wiki page):

- $det(\mathbf{I}) = 1$
- $det(A^T) = det(A)$ (transposing the matrix does not affect the determinant)
- $det(A^{-1}) = \frac{1}{det(A)}$ (the determinant of the inverse is the inverse of the determinant)
- For A, B square matrices of equal size, $det(AB) = det(A)det(B)$
- $det(cA) = c^n \, det(A)$ for any scalar c
- If A is triangular (so has all zeros in the upper or lower triangle) then $det(A) = \prod_{i=1}^{n} A_{ii}$.

An *eigenvector* of the square matrix $\mathbf{A}$ is a non-zero vector $\mathbf{e}$ such that $\mathbf{Ae} = \lambda\mathbf{e}$ for some scalar λ. λ is known as the *eigenvalue* of $\mathbf{A}$ corresponding to $\mathbf{e}$. Note that λ may be 0. So the effect of multiplying e by A is simply to scale e by the corresponding scalar λ.

The determinant can be used to find the eigenvalues of $\mathbf{A}$: they are the roots of the *characteristic polynomial* $p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}_n)$ where $\mathbf{I}_n$ is the identity matrix.

Example: Find the eigenvalues of $\mathbf{A} = \begin{pmatrix} 3 & 5 \\ 1 & -1 \end{pmatrix}$.

Solution: We need to solve $p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}_2) = 0$.

$$\begin{aligned} |\mathbf{A} - \lambda\mathbf{I}_2| &= \left| \begin{pmatrix} 3 & 5 \\ 1 & -1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| \\ &= \begin{vmatrix} 3-\lambda & 5 \\ 1 & -1-\lambda \end{vmatrix} \\ &= (3-\lambda)(-1-\lambda) - 5 \\ &= -\lambda^2 - 2\lambda - 8 \\ &= (\lambda+2)(\lambda-4) \end{aligned}$$

which is zero when $\lambda = 4$ or $\lambda = -2$. So the eigenvalues of $\mathbf{A}$ are $\lambda = 4$ and $\lambda = -2$. □

Example: Find the eigenvector of $\mathbf{A} = \begin{pmatrix} 3 & 5 \\ 1 & -1 \end{pmatrix}$ corresponding to the eigenvalue $\lambda = -2$.

Solution: The eigenvector $\mathbf{e}$ corresponding to $\lambda = -2$ satisfies the equation $\mathbf{Ae} = -2\mathbf{e}$. That is,

$$\begin{pmatrix} 3 & 5 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = -2 \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}.$$

This is the system of linear equations

$$\begin{aligned} 3e_1 + 5e_2 &= -2e_1, & (1) \\ e_1 - e_2 &= -2e_2. & (2) \end{aligned}$$

Rearranging either equation, we get $e_1 = -e_2$, so both equations are the same. We thus fix $e_1 = 1$ and the eigenvector associated with $\lambda = -2$ is $\mathbf{e} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Notice that the choice to fix $e_1 = 1$ was arbitrary. We could choose any value so, strictly, $\mathbf{e} = c \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ for any $c \neq 0$. Often, c is chosen so that $\mathbf{e}$ is normalised (see below). In this case, choose $c = 1/\sqrt{2}$ to normalise $\mathbf{e}$. □

Vectors a and b are *orthogonal* if the dot product $a^T b = 0$. Orthogonal generalises the of the idea of the perpendicular. In particular, a set of vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is *mutually orthogonal* if each pair of vectors e_i, e_j is orthogonal for $i \neq j$.

A vector $\mathbf{e}_i$ is normalised if $\mathbf{e}_i^T e_i = 1$.

A set of vectors that is mutually orthogonal and has each vector normalise is called *orthonormal*.

Any symmetric, square matrix $\mathbf{A}$ of size n has exactly n eigenvectors that are mutually orthogonal.

Any square matrix A of size n that has n mutually orthogonal eigenvectors can be represented via the *eigenvector representation* as follows:

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \underbrace{\mathbf{e}_i \mathbf{e}_i^{\mathsf{T}}}_{\mathbf{U}_i}$$

where $\mathbf{U}_i = \mathbf{e}_i \mathbf{e}_i^{\mathsf{T}}$ is an $n \times n$ matrix.

The *Range*, range($\mathbf{A}$), or span of an $m \times n$ matrix $\mathbf{A}$ is the set of vectors $\mathbf{y} \in \mathbb{R}^m$ such that $\mathbf{y} = \mathbf{Ax}$ for some $\mathbf{x} \in \mathbb{R}^n$. The range is also referred to as the *column space* of $\mathbf{A}$ as it is the space of all linear combinations of the columns of $\mathbf{A}$.

The *Nullspace*, null($\mathbf{A}$), of an $m \times n$ matrix $\mathbf{A}$ is the set of vectors $\mathbf{x} \in \mathbb{R}^n$, such that $\mathbf{Ax} = \mathbf{0} \in \mathbb{R}^m$

The *Rank*, rank($\mathbf{A}$), of an $m \times n$ matrix $\mathbf{A}$ is the dimension of the range of $\mathbf{A}$ or of the column space of $\mathbf{A}$. rank($\mathbf{A}$) $\leq \min\{m, n\}$.