MultiGrid Calculus

Author: Han de Bruijn (1999)
Latest revision: 2004/11/03

The well-known Newton-Rhapson algorithm is used as a starting point for still another method for inverting tri-diagonal matrices. It is shown that this method is closely related to MultiGrid algorithms. The notion of Persistent Properties is developed. The quotient of the off-diagonal matrix coefficients proves to be an exponential function of the grid spacing. The behaviour of a product of the matrix coefficients can be understood in full detail, with help of a Connection to Trigonometry in a *dangerous* domain and a Hyperbolic Connection in a *safe* domain. The safe domain is quite distinct from the dangerous one. It is shown that all safe solutions form a sampling of the analytical solutions of the second order linear ODE (Ordinary Differential Equation). But it is demanded that the *discriminant* of this ODE is positive or zero. The matrix coefficients can be expressed in the coefficients of the ODE and the grid spacing.

Newton-Rhapson algorithm

The Newton-Rhapson method is a numerical algorithm for finding zeros p of a function $f(x)$. The gist of the method is to draw successive tangent lines and determine where these lines intersect the x-axis (see figure on next page):

$$y - f(p_n) = f'(p_n).(x - p_n) \quad \text{where} \quad y = 0 \quad \text{and} \quad x = p_{n+1} \quad \Longrightarrow$$

$$p_{n+1} = p_n - \frac{f(p_n)}{f'(p_n)}$$

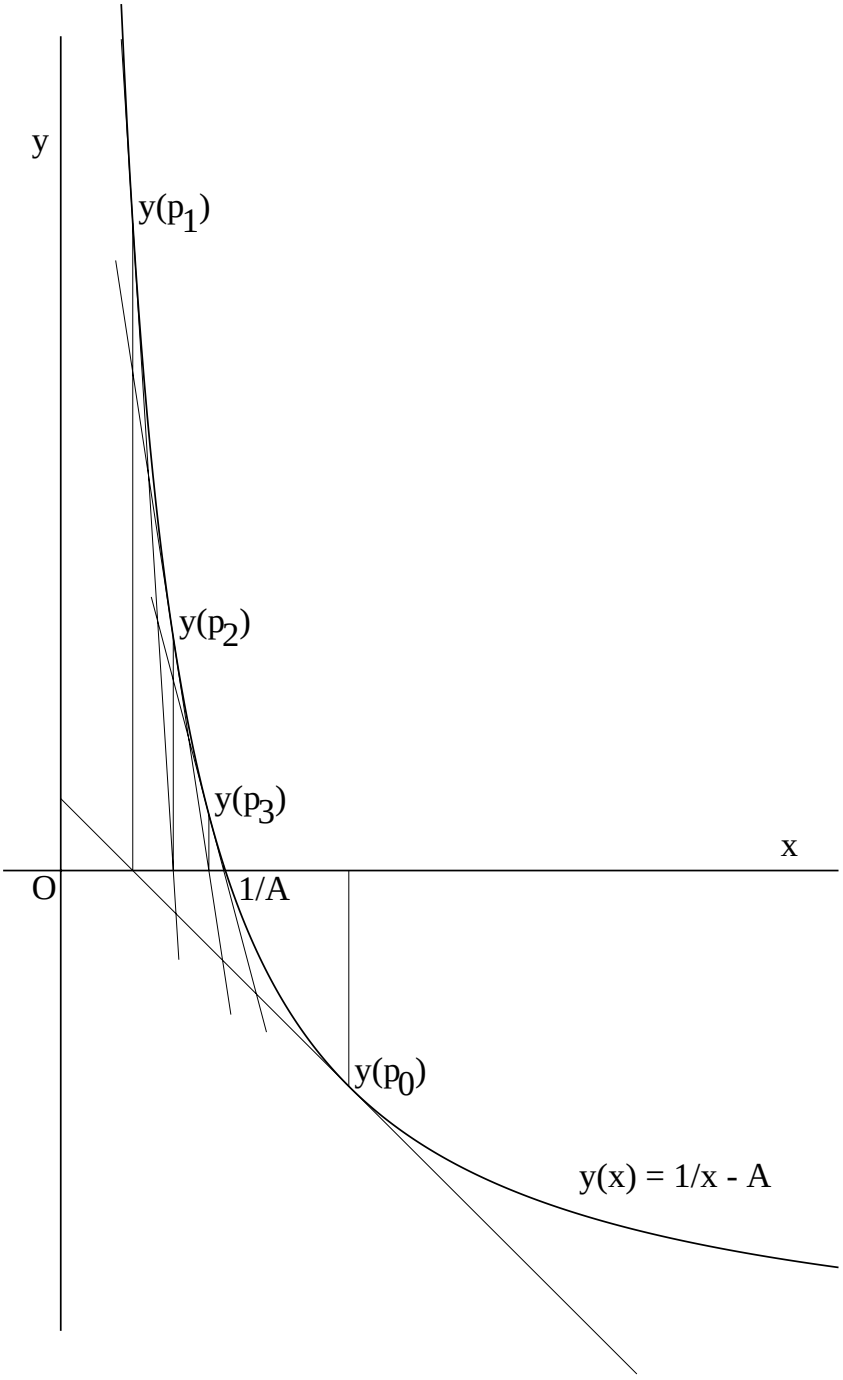
Thereby assuming that the whole process will be convergent.
The method can be used for performing a division without actually performing a division. (I found this in a reader about Numerical Analysis.) The equation to be searched for its roots, in this case, is given by:

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

Substitute $f(x) = 1/x - a$ in the above algorithm. Resulting in:

$$p_{n+1} = p_n - \frac{1/p_n - a}{-1.1/p_n^2} = p_n - (-p_n + a.p_n^2) \quad \Longrightarrow \quad p_{n+1} = 2.p_n - a.p_n^2$$

It is well known that the Newton-Rhapson method, if it converges, then it does so quadratically, meaning that the inverse $1/a$ can be found rather quickly. The algorithm has been used in the old days, on computers which had no floating point division instruction available.


y
y(p1)
y(p2)
y(p3)
x
O
1/A
y(p0)
y(x) = 1/x - A

.

Thus by employing the Newton-Rhapson algorithm, the inverse of a number can be found with quadratic speed, by performing solely additions and multiplications. Armed with this knowledge, let's make the transition from numbers to matrices. Determining the inverse of a matrix, filled with many numbers, seems to be much more like a challenge anyway.
Let the iterative process for matrices be defined by:

$$P_0 = I \quad \text{and} \quad P_{n+1} = 2.P_n - A.P_n.P_n$$

Here I is the unity matrix, A is the matrix to be inverted and P_n are the successive iterands, which should converge to the inverse matrix A^{-1}.
An "initial guess" or "preconditioner" T may be chosen instead of the unity matrix I. Such a preconditioner T may be equal to the inverse of the main diagonal of A, if things are to be kept simple. This effectively means that each row of A (and each element of the right hand side b) is divided by the accompanying main diagonal element. The result is a system of equations which is commonly called *normed*. In a normed system of equations, all elements of the main diagonal are equal to 1. With other words: the main diagonal is equal to the identity matrix. It is noted that an eventual symmetry of A will be destroyed by carrying out this process of *normalization*. However, it will be assumed below, for the case of simplicity, that the starting point is just the unity matrix. Hence $T = I$.

Theorem:

$$\text{Let} \quad M = (I - A) \quad \text{or} \quad A = (I - M) \quad \text{then:} \quad P_n = (I - M^{2^n}).A^{-1}$$

Proof by induction:

$$\begin{aligned}
P_0 &= I = A.A^{-1} = (I - M).A^{-1} \\
P_{n+1} &= \left(I - M^{2^{n+1}}\right).A^{-1} \\
&= \left(I - M^{2^n.2}\right).A^{-1} \\
&= \left[I - (M^{2^n})^2\right].A^{-1} \\
&= \left(I - M^{2^n}\right).\left(I + M^{2^n}\right).A^{-1} \\
&= \text{because all matrices are mutually commutative:} \\
&= \left(I - M^{2^n}\right).A^{-1}.\left[2.I - A.\left(I - M^{2^n}\right).A^{-1}\right] \\
&= P_n.(2.I - A.P_n) = 2.P_n - A.P_n.P_n
\end{aligned}$$

Let $m = 2^n$, then:

$$P_n = (I - M^m).(I - M)^{-1}$$

For numbers, this would be the sum of a Geometric Series:

$$(I - M^m).(I - M)^{-1} = (I + M + M^2 + M^3 + M^4 + M^5 + M^6 + ... + M^{m-1})$$

For matrices, this Geometric Series turns out to be equivalent to an iterative "incremental Jacobi" solution method, as will be explained in Appendix I. But there also does exist a *product* of terms, called the Euler expansion:

$$(I - M^m).(I - M)^{-1} = (I + M^{m/2}).(I - M^{m/2}).(I - M)^{-1} =$$
$$(I + M^{m/2}).(I + M^{m/4}).(I - M^{m/4}).(I - M)^{-1} =$$
$$(I + M^{m/2}).(I + M^{m/4}). ... (I + M^2).(I + M).(I - M).(I - M)^{-1} =$$
$$(I + M^{m/2}).(I + M^{m/4}). ... (I + M^8).(I + M^4).(I + M^2).(I + M)$$

Let the system of equations to be solved be given by $A.w = b$. Using the above sequences, we can write:

$$P_n = (I - M^{2^n}).A^{-1} \quad \Longrightarrow \quad A^{-1}b = (I - M^{2^n})^{-1}.P_n b$$

$$= (I - M^{2^n})^{-1}.(I + M^{m/2}).(I + M^{m/4}). ... (I + M^8).(I + M^4).(I + M^2).(I + M)b$$

Another way of looking at this is the following:

$$(I - M)^{-1} = (I - M)^{-1}.(I + M)^{-1}.(I + M) = (I - M^2)^{-1}.(I + M) =$$
$$(I - M^2)^{-1}.(I + M^2)^{-1}.(I + M^2).(I + M) = (I - M^4)^{-1}.(I + M^2).(I + M)$$

With other words:

$$w = (I - M)^{-1}.b = (I - M)^{-1}.(I + M)^{-1}.(I + M).b = (I - M^2)^{-1}.(I + M).b$$

Define $b := (I + M).b$ and $M := M^2$. Then again:

$$w = (I - M)^{-1}.b = (I - M)^{-1}.(I + M)^{-1}.(I + M).b = (I - M^2)^{-1}.(I + M).b$$

And this process can be repeated until we find a way to determine $(I - M)^{-1}$, preferrably without continuing the iterations ad infinitum. However, is there any reason to believe that M^2 is more amenable, one way or another, than M?

2×2 matrix

The method outlined so far will be applied to the simplest non-trivial example, which is a matrix of rank 2. The system of equations to be solved is:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} p \\ q \end{bmatrix}$$

Assume that $a \neq 0$ and $d \neq 0$. At first, the equations will be normed:

$$\begin{bmatrix} 1 & b/a \\ c/d & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} p/a \\ q/d \end{bmatrix}$$

The matrix M is formed by subtracting from the identity matrix:

$$M = \begin{bmatrix} 0 & -b/a \\ -c/d & 0 \end{bmatrix}$$

Hence:

$$I + M = \begin{bmatrix} 1 & -b/a \\ -c/d & 1 \end{bmatrix}$$

And:

$$M^2 = \begin{bmatrix} 0 & -b/a \\ -c/d & 0 \end{bmatrix} . \begin{bmatrix} 0 & -b/a \\ -c/d & 0 \end{bmatrix} = \begin{bmatrix} b/a.c/d & 0 \\ 0 & c/d.b/a \end{bmatrix}$$

Hence:

$$I - M^2 = \begin{bmatrix} 1 - b/a.c/d & 0 \\ 0 & 1 - c/d.b/a \end{bmatrix}$$

Nothing is simpler than determining the inverse of a diagonal matrix:

$$(I - M^2)^{-1} = \begin{bmatrix} 1/(1 - b/a.c/d) & 0 \\ 0 & 1/(1 - c/d.b/a) \end{bmatrix}$$

Hence:

$$\frac{I + M}{I - M^2} = \begin{bmatrix} 1/(1 - b/a.c/d) & 0 \\ 0 & 1/(1 - c/d.b/a) \end{bmatrix} . \begin{bmatrix} 1 & -b/a \\ -c/d & 1 \end{bmatrix} =$$

$$\frac{1}{1 - b/a.c/d} . \begin{bmatrix} 1 & -b/a \\ -c/d & 1 \end{bmatrix}$$

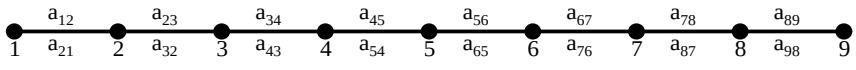
It is easily recognized that this is *exactly* the inverse of:

$$\begin{bmatrix} 1 & b/a \\ c/d & 1 \end{bmatrix}$$

It is concluded that any system of 2 equations with 2 unknowns is solved *exactly* by the Newton-Rhapson method (provided that the main diagonal elements are non-zero and the matrix is non-singular).

MultiGrid

A possible implementation of the Newton-Rhapson method for linear equations is triggered by the following observation. Consider a one-dimensional grid:



Then, in almost all cases, only *adjacent* grid-points will be connected by certain

(matrix) coefficients. This gives rise to a tri-diagonal system of equations. Let's assume beforehand that the equations are *always* normed (which means that the elements of the main diagonal must be non-zero from the start). Thus, without loss of generality, we may assume:

$$A = \begin{bmatrix} 1 & a_{12} & & & & & & & \\ a_{21} & 1 & a_{23} & & & & & & \\ & a_{32} & 1 & a_{34} & & & & & \\ & & a_{43} & 1 & a_{45} & & & & \\ & & & a_{54} & 1 & a_{56} & & & \\ & & & & a_{65} & 1 & a_{67} & & \\ & & & & & a_{76} & 1 & a_{78} & \\ & & & & & & a_{87} & 1 & a_{89} \\ & & & & & & & a_{98} & 1 \end{bmatrix}$$

The matrix M is then formed by $M = I - A$:

$$\begin{bmatrix} 0 & -a_{12} & & & & & & & \\ -a_{21} & 0 & -a_{23} & & & & & & \\ & -a_{32} & 0 & -a_{34} & & & & & \\ & & -a_{43} & 0 & -a_{45} & & & & \\ & & & -a_{54} & 0 & -a_{56} & & & \\ & & & & -a_{65} & 0 & -a_{67} & & \\ & & & & & -a_{76} & 0 & -a_{78} & \\ & & & & & & -a_{87} & 0 & -a_{89} \\ & & & & & & & -a_{98} & 0 \end{bmatrix}$$

And the matrix $I - M^2$ can be computed. It turns out to be of the form:

$$\begin{bmatrix} a_{11} & 0 & a_{13} & & & & & & \\ 0 & a_{22} & 0 & a_{24} & & & & & \\ a_{31} & 0 & a_{33} & 0 & a_{35} & & & & \\ & a_{42} & 0 & a_{44} & 0 & a_{46} & & & \\ & & a_{53} & 0 & a_{55} & 0 & a_{57} & & \\ & & & a_{64} & 0 & a_{66} & 0 & a_{68} & \\ & & & & a_{75} & 0 & a_{77} & 0 & a_{79} \\ & & & & & a_{86} & 0 & a_{88} & 0 \\ & & & & & & a_{97} & 0 & a_{99} \end{bmatrix}$$

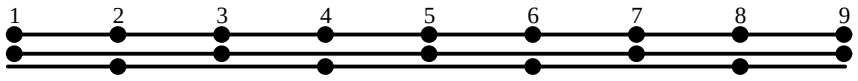
It is observed now that all even and odd indices have become *uncoupled*. The variables can be permuted and the equations be rearranged accordingly:

$$\left[\begin{array}{ccccc|cccc}
a_{11} & a_{13} & & & & & & & \\
a_{31} & a_{33} & a_{35} & & & & & & \\
 & a_{53} & a_{55} & a_{57} & & & & & \\
 & & a_{75} & a_{77} & a_{79} & & & & \\
 & & & a_{97} & a_{99} & & & & \\
\hline
 & & & & & a_{22} & a_{24} & & \\
 & & & & & a_{42} & a_{44} & a_{46} & \\
 & & & & & & a_{64} & a_{66} & a_{68} \\
 & & & & & & & a_{86} & a_{88}
\end{array}\right]$$

The new equations system is of the form:

$$\left[\begin{array}{cc} A_1 & 0 \\ 0 & A_2 \end{array}\right]$$

Block A_1 corresponds with a restriction of the grid to odd-numbered points, block A_2 corresponds with a restriction of the grid to even-numbered points, as depicted in the figure below. Such a configuration of coarsened grids will be called a **MultiGrid** in the sequel.



Once having a *blocked* system of equations, each block can be solved quite independently of the other. Thus in fact we have *two* independent systems of equations, A_1 and A_2, each of them having the same structure as the original system A. These equations shall be *renormalized* at each step, thus assuring that also the new matrices have a diagonal equal to the identity. Probably we can apply the same procedure again and again, thereby *reducing* any tri-diagonal matrix into ever smaller blocks, which will finally become just single numbers along the main diagonal. After normalization of the latter we should be finished. A well documented implementation of the Newton-Rhapson MultiGrid method, programmed in (Turbo) Pascal, is found in Appendix II.

Direct Solver

We will take a closer look at the procedure which forms the matrix $(1-M^2)$ out of the tri-diagonal matrix A. It is assumed that the tri-diagonal system matrix A has been (re)normalized. Then the general form of such a matrix is:

$$\begin{bmatrix} \cdot & \cdot & \cdot & & & & \\ & a_{43}/a_{44} & 1 & a_{45}/a_{44} & & & \\ & & a_{54}/a_{55} & 1 & a_{56}/a_{55} & & \\ & & & a_{65}/a_{66} & 1 & a_{67}/a_{66} & \\ & & & & \cdot & \cdot & \cdot \end{bmatrix} = A = I - M$$

The matrix M^2 is calculated. Concentrate on row (5) and columns $(4,5,6)$:

$$\begin{bmatrix} \cdot & \cdot & & & \\ & \cdot & \cdot & \cdot & \\ & & -\frac{a_{54}}{a_{55}} & 0 & -\frac{a_{56}}{a_{55}} \\ & & & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot & \cdot & & \\ & -\frac{a_{43}}{a_{44}} & 0 & -\frac{a_{45}}{a_{44}} & \\ & & -\frac{a_{54}}{a_{55}} & 0 & -\frac{a_{56}}{a_{55}} \\ & & & -\frac{a_{65}}{a_{66}} & 0 \\ & & & & \cdot \end{bmatrix}$$

$$= \begin{bmatrix} \cdot & \cdot & \cdot & & \\ & \cdot & & \cdot & \\ & \frac{a_{54}}{a_{55}}\frac{a_{43}}{a_{44}} & 0 & \frac{a_{54}}{a_{55}}\frac{a_{45}}{a_{44}} + \frac{a_{56}}{a_{55}}\frac{a_{65}}{a_{66}} & 0 \\ & & & \cdot & \cdot \end{bmatrix}$$

Completing further one step of the Newton-Rhapson procedure results in:

$$I - M^2 = \begin{bmatrix} \cdot & \cdot & \cdot & & & \\ & \cdot & & \cdot & & \\ & -\frac{a_{54}}{a_{55}}\frac{a_{43}}{a_{44}} & 0 & 1 - \frac{a_{54}}{a_{55}}\frac{a_{45}}{a_{44}} - \frac{a_{56}}{a_{55}}\frac{a_{65}}{a_{66}} & 0 & -\frac{a_{56}}{a_{55}}\frac{a_{67}}{a_{66}} \\ & & & \cdot & \cdot & \\ & & & & \cdot & \cdot \end{bmatrix}$$

It is shown in the following sequence that a Newton Rhapson step for coarsening the grid is equivalent with an Gaussian elimination step for even grid points. Hence the method as a whole is equivalent with a *direct solution method* for the linear equations system. This explains why an exact solution is found in the first place.

$$\begin{bmatrix} \cdot & \cdot & \cdot & & & & \\ & a_{43}/a_{44} & 1 & a_{45}/a_{44} & & & \\ & & a_{54}/a_{55} & 1 & a_{56}/a_{55} & & \\ & & & a_{65}/a_{66} & 1 & a_{67}/a_{66} & \\ & & & & \cdot & \cdot & \cdot \end{bmatrix} = A$$

Use the first and the last row for pivoting the middle row. This is equivalent with:

$$T_4 = -a_{43}/a_{44}.T_3 - a_{45}/a_{44}.T_5$$
$$T_6 = -a_{65}/a_{66}.T_5 - a_{67}/a_{66}.T_7$$

Substitute in the row with index (5):

$$a_{54}/a_{55}.T_4 + 1.T_5 + a_{56}/a_{55}.T_6 = 0 \quad \Longrightarrow$$

$$\frac{a_{54}}{a_{55}}\left[-\frac{a_{43}}{a_{44}}T_3 - \frac{a_{45}}{a_{44}}T_5\right] + T_5 + \frac{a_{56}}{a_{55}}\left[-\frac{a_{65}}{a_{66}}T_5 - \frac{a_{67}}{a_{66}}T_7\right]$$
$$= -\frac{a_{54}}{a_{55}}\frac{a_{43}}{a_{44}}T_3 + \left[1 - \frac{a_{54}}{a_{55}}\frac{a_{45}}{a_{44}} - \frac{a_{56}}{a_{55}}\frac{a_{65}}{a_{66}}\right]T_5 - \frac{a_{56}}{a_{55}}\frac{a_{67}}{a_{66}}T_7 = 0$$

When written in matrix form, the abovementioned equivalence becomes evident:

$$\begin{bmatrix} \cdot & \cdot & \cdot & & & & & \\ & & \cdot & & \cdot & & & \\ & -\frac{a_{54}}{a_{55}}\frac{a_{43}}{a_{44}} & 0 & 1 - \frac{a_{54}}{a_{55}}\frac{a_{45}}{a_{44}} - \frac{a_{56}}{a_{55}}\frac{a_{65}}{a_{66}} & 0 & -\frac{a_{56}}{a_{55}}\frac{a_{67}}{a_{66}} & & \\ & & & \cdot & \cdot & \cdot & & \\ & & & & \cdot & \cdot & \cdot & \end{bmatrix} \begin{bmatrix} \cdot \\ \cdot \\ T_3 \\ T_4 \\ T_5 \\ T_6 \\ T_7 \\ \cdot \end{bmatrix}$$

Persistent Properties

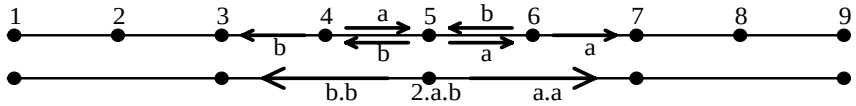
Suppose we have a uniform grid and let $j > i$. Then all coefficients a_{ij} may be assumed to be equal to each other. The same is true for all coefficients a_{ji}. Let $a_{ij} = -a$ and $a_{ji} = -b$. Also assume that the system matrix S has been (re)normalized. Then the general form of such a matrix is:

$$\begin{bmatrix} . & . & . & & & & \\ & -b & 1 & -a & & & \\ & & -b & 1 & -a & & \\ & & & -b & 1 & -a & \\ & & & & . & . & . \end{bmatrix} = S = I - M$$

The matrix M^2 is calculated:

$$\begin{bmatrix} . & . & . & & & & \\ & b^2 & 2.a.b & a^2 & & & \\ & & b^2 & 2.a.b & a^2 & & \\ & & & b^2 & 2.a.b & a^2 & \\ & & & & . & . & . \end{bmatrix} = M^2$$

It is seen that, at a coarser grid, the coefficients are obtained by travelling two steps in any direction on the refined grid, multiplying the accompanying matrix-coefficients with each other and adding together the results of each of the possible paths. (This vaguely reminds to a piece of QuantumElectroDynamics.) Hence, travelling from (5) to (4) gives a contribution $b.b$, travelling from (5) to (7) gives a contribution $a.a$ and travelling from (5) to (5) two times back and forth gives a contribution $a.b + b.a$:



Completing further one step of the Newton-Rhapson procedure results in:

$$\begin{bmatrix} . & . & . & & & & \\ & -b^2 & 1-2.a.b & -a^2 & & & \\ & & -b^2 & 1-2.a.b & -a^2 & & \\ & & & -b^2 & 1-2.a.b & -a^2 & \\ & & & & . & . & . \end{bmatrix} = I - M^2 = S'$$

Where S' is a blocked matrix, corresponding with 2 coarser grids. Now it is sensible to require that $(I - M^2)$, after re-normalization, has properties which are more or less alike those of $(I - M)$. Essentially the same kind of discretization scheme should be used at *any* of the coarser or refined grids. It's useful to give a name to the phenomenon. If certain properties of a scheme remain the same

at any of the MultiGrids, then such Properties will be called **Persistent**.
After renormalization, the off-diagonal coefficients become:

$$a' = \frac{a^2}{1 - 2.a.b} \quad \text{and} \quad b' = \frac{b^2}{1 - 2.a.b}$$

Under the obvious assumption that $1 - 2.a.b \neq 0$.
And we are ready to find our first *persistent* properties, though maybe they are somewhat trivial:

$$a = 0 \iff a' = 0 \qquad \text{and} \qquad b = 0 \iff b' = 0$$

So far so good. Let's consider cases where $a \neq 0$ and $b \neq 0$. From the coarsening formulas for a and b, it can then be inferred immediately that:

$$\frac{a'}{b'} = \frac{a^2}{b^2} > 0$$

While multiplying a' and b' gives:

$$a'.b' = \frac{a^2.b^2}{(1 - 2.a.b)^2} > 0$$

Four different cases may be distinguished:

$$\begin{array}{llll} a > 0 & \text{and} & b < 0 & \qquad a < 0 \quad \text{and} \quad b > 0 \\ a < 0 & \text{and} & b < 0 & \qquad a > 0 \quad \text{and} \quad b > 0 \end{array}$$

The formulas for the quotient and the product show that only a *positive* product and a *positive* quotient of the off-diagonal coefficients can be actually persistent through coarser multigrids. Hence combinations $a > 0, b < 0$ and $a < 0, b > 0$ are already out of the question. Thus, the only properties which may be invariant for grid coarsening are: both coefficients negative or both coefficients positive. We are ready now to infer one of the inverse functions, which is applicable to grid refinement instead of grid coarsening:

$$\frac{a'}{b'} = \frac{a^2}{b^2} \iff \frac{a}{b} = \sqrt{\frac{a'}{b'}}$$

Substituting $a = \sqrt{a'/b'}.b$ into $b' = b^2/(1 - 2.a.b)$ leads to an equation from which b can be solved:

$$b' = b^2/(1 - 2.a.b) \implies b' - 2.a.b.b' = b^2 \implies b' - 2.\sqrt{a'/b'}.b.b.b' = b^2$$

$$\implies b' = 2.\sqrt{a'.b'}.b^2 + b^2 \implies b' = b^2.(1 + 2.\sqrt{a'.b'}) > 0$$

Almost the same can be done for a:

$$a' = a^2.(1 + 2.\sqrt{a'.b'}) > 0$$

Herewith also the condition $a < 0, b < 0$ is ruled out as a possible persistent property, leaving $a > 0, b > 0$ as the only possibility left. We conclude that the following is a *necessary condition for persistence*, given $-a$ and $-b$ as the off-diagonal elements in a (normed) tri-diagonal system, while admitting again the special cases $a = 0$ and/or $b = 0$:

$$a \geq 0 \quad \text{and} \quad b \geq 0$$

The well known and commonly accepted "rule of positive coefficients" (Patankar 1980) is found back herewith.
The inverse of mesh coarsening, grid refinement, seems to be at least equally important as the former. Given the discretization on a coarser grid, how can we obtain then the discretization on a finer grid? It is clear that grid refinement, in general, can be quite ambiguous. That's one reason why, probably, we shall have to restrict ourselves to uniform grids.
Inverse functions can be derived for the off-diagonal coefficients as such:

$$a' = a^2/(1 - 2.a.b) \quad \Longleftrightarrow \quad a = \sqrt{\frac{a'}{1 + 2.\sqrt{a'.b'}}}$$

$$b' = b^2/(1 - 2.a.b) \quad \Longleftrightarrow \quad b = \sqrt{\frac{b'}{1 + 2.\sqrt{a'.b'}}}$$

Would it be possible that the coefficients on the fine grid are not different at all from those on the coarse grid? The answer on this question is affirmative:

$$\begin{aligned} a = a^2/(1 - 2.a.b) \quad &\Longrightarrow \quad 1 - 2.a.b = a \\ b = b^2/(1 - 2.a.b) \quad &\Longrightarrow \quad 1 - 2.a.b = b \end{aligned}$$

We have divided by a and b and it should be noted that $a = 0$ and/or $b = 0$ is also a solution. In the asymmetric case we may assume $a > 0$, $b = 0$ or vice versa, resulting in:

$$\begin{aligned} b = 0 \quad &\Longrightarrow \quad a = a^2 \quad \Longrightarrow \quad a = 1 \\ a = 0 \quad &\Longrightarrow \quad b = b^2 \quad \Longrightarrow \quad b = 1 \end{aligned}$$

For all other cases, it can be concluded, in general, that $a = b$. Substitute this in one of the equations:

$$1 - 2.a.a = a \quad \Longrightarrow \quad a^2 + \frac{1}{2}a - \frac{1}{2} = (a + 1)(a - \frac{1}{2}) = 0 \quad \Longrightarrow \quad a = \frac{1}{2}$$

The other solution $a = -1$ is not persistent because of the rule of positive coefficients. Thus we have *four* cases for which the (bulk) coefficients are the

same (and of course *persistent*) on all of the MultiGrids:

$$\begin{aligned} a = b = 0 &\qquad \text{Identity matrix} \\ a = 0, b = 1 &\qquad \text{Lower diagonal matrix} \\ a = 1, b = 0 &\qquad \text{Upper diagonal matrix} \\ a = b = \frac{1}{2} &\qquad \text{Symmetric matrix} \end{aligned}$$

Now take the *sum* of the off-diagonal coefficients:

$$a' + b' = \frac{a^2 + b^2}{1 - 2.a.b} \quad \Longleftrightarrow \quad a + b = \frac{\sqrt{a'} + \sqrt{b'}}{\sqrt{1 + 2.\sqrt{a'.b'}}}$$

Rewrite the formula for grid coarsening as follows:

$$a' + b' = \frac{(a+b)^2 - 2.a.b}{1 - 2.a.b}$$

Then it is clear that all of the following properties are *persistent* at coarsened as well as refined grids:

$$\begin{aligned} a + b < 1 &\quad \Longleftrightarrow \quad a' + b' < 1 \\ a + b = 1 &\quad \Longleftrightarrow \quad a' + b' = 1 \\ a + b > 1 &\quad \Longleftrightarrow \quad a' + b' > 1 \end{aligned}$$

One subject we have barely touched is the sign of the denominator, which arises as soon as the matrix $I - M^2$ is normed in the Newton-Rhapson procedure:

$$1 - 2.a.b \neq 0$$

Using the rule of positive coefficients and the coarsening formulas for a' and b', we can even write:

$$1 - 2.a.b > 0 \quad \Longrightarrow \quad a.b < \frac{1}{2}$$

It can be shown that this property is *not* a persistent one:

$$0 < 1 - 2.a'.b' = 1 - \frac{a^2.b^2}{(1 - 2.a.b)^2} = \frac{1 - 4.(a.b) + 4.(a.b)^2 - (a.b)^2}{(1 - 2.a.b)^2} > 0 \quad \Longleftrightarrow$$

$$3\left[(a.b)^2 - 4/3.(a.b) + 1/3\right] > 0 \quad \Longleftrightarrow \quad [a.b - 1]\,[a.b - 1/3] > 0$$

The accompanying function is a parabola:

$$y = \left(x - \frac{2}{3}\right)^2 - \frac{4}{9} + \frac{3}{9}$$

Which has a minimum for $(x, y) = (2/3, -1/9)$. While the left hand side of the preceding expression is positive for $a'.b' < 1/2$, the right hand side is positive for $a.b < 1/3$ or $a.b > 1$. The combined outcome of the two is: $a.b < 1/3 < 1/2$. Further grid coarsening gives rise to an equation of the 4th order, at least. So we have no idea (yet) how these findings may eventually lead to a persistent property for the denominators.

Quotient Function

Suppose the off-diagonal coefficients at the fine grid are a and b, and the off-diagonal coefficients at the coarser grid are a' and b', then we know from the chapter about "Persistent Schemes" that the following relationships hold:

$$\frac{a'}{b'} = \left(\frac{a}{b}\right)^2 \quad \text{and} \quad \frac{a}{b} = \sqrt{\frac{a'}{b'}}$$

It is inferred that the relative magnitude of the coefficients a and b is *persistent* through grid coarsening and grid refinement:

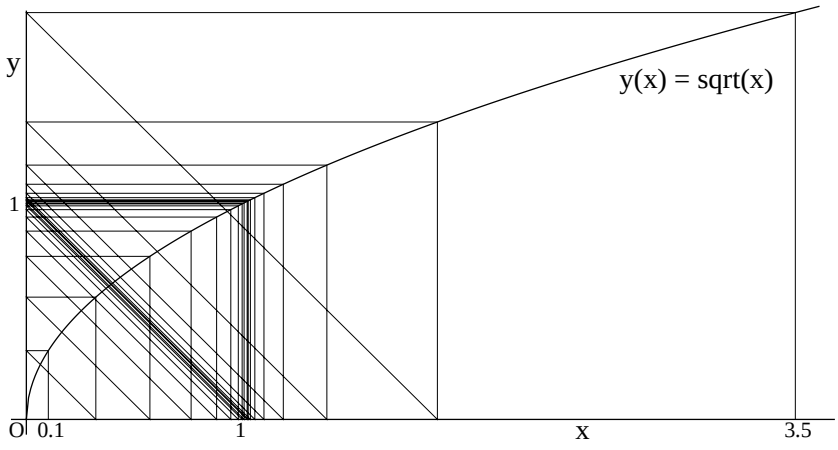
$$\begin{aligned} a < b \quad &\Longleftrightarrow \quad a' < b' \\ a = b \quad &\Longleftrightarrow \quad a' = b' \\ a > b \quad &\Longleftrightarrow \quad a' > b' \end{aligned}$$

With grid coarsening, the fraction a/b is subject to the function $x' = x^2$. The outcome of which can be subject to another grid coarsening $x'' = x'^2$, to another grid coarsening $x''' = x''^2$ and so on and so forth. Resulting in an expression:

$$\left(\left(\left(\left(...x^2\right)^2\right)^2...\right)^2\right)^2 = \lim_{N\to\infty} x^{2^N} = 0\,,1\,,\,\infty \quad \text{for} \;\; x<1, x=1, x>1$$

With grid refinement, the fraction a/b is subject to the function $x' = \sqrt{x}$. The outcome of which can be subject to another grid refinement $x'' = \sqrt{x'}$, to another grid refinement $x''' = \sqrt{x''}$ and so on and so forth. Resulting in an expression:

$$\sqrt{\sqrt{\sqrt{\sqrt{\sqrt{...\sqrt{x}}}}}} = \lim_{N\to\infty} \sqrt[2^N]{x} = 1 \quad \text{for} \;\; x>0$$



It is kind of a custom to say, nowadays, that the function $y = \sqrt{x}$ has an *attractor* for $x = y = 1$. The practical meaning is that, at infinitely refined grids, a and b will become equal: $a = b$ for grid $\rightarrow \infty$-dense. (Provided that $a \neq 0$ and $b \neq 0$; for if else they will remain zero forever.) The point $(1, 1)$ itself is a stable stationary point: any point in the neighbourhood will become more and more "equal" to it and the point $(1, 1)$ itself does not change with further iterations.
It is seen, on the contrary, that the point $x = y = 1$ of the inverse function $y = x^2$ is not an attractor, but more like a *repeller* instead. Values smaller than 1 are pushed towards zero, while values greater that 1 are pushed towards infinity. The point $x = y = 1$ itself is a stationary point, but it is highly unstable.
If a grid becomes 2 times coarser, then the new quotient of a and b will be related to the old one by: $(a/b) := (a/b)^2$. If a grid becomes $1/2$ times coarser, then the new quotient of the off-diagonal elements will be related to the old one by: $(a/b) := (a/b)^{1/2}$. We could also have written:

$$\frac{a}{b}(2.dx) = \left[\frac{a}{b}(dx)\right]^2 \quad \text{and} \quad \frac{a}{b}(\frac{1}{2}dx) = \left[\frac{a}{b}(dx)\right]^{1/2}$$

We seek to generalize these results, where a relationship as the following comes into mind:

$$\frac{b}{a}\,(p.dx) = \left[\frac{b}{a}\,(dx)\right]^p \quad \text{for any real number } p$$

An interpretation for negative numbers p can be obtained as follows. Consider a one-dimensional grid and suppose that it is traversed in the reverse direction. Thus, instead of numbering grid points from the left to the right, they are numbered from the right to the left. By this "inverse" transformation, any matrix coefficient a_{ij} with $i > j$ will be mapped upon a coefficient a_{ji} and any coefficient a_{ji} will be mapped upon a matrix coefficient a_{ij}. Using uniform meshes, we see that coefficients a are transformed into coefficients b and coefficients b are transformed into coefficients a. This means that fractions a/b will be transformed into fractions b/a and vice versa. But travelling the grid in reverse order also means that we are using increments $(-dx)$ instead of increments $(+dx)$. Therefore we can write:

$$\frac{a}{b}(-dx) = \frac{b}{a}(dx) = \left[\frac{a}{b}(dx)\right]^{-1} \quad \text{and} \quad \frac{b}{a}(-dx) = \frac{a}{b}(dx) = \left[\frac{b}{a}(dx)\right]^{-1}$$

For $dx = 0$ it follows that $-dx = +dx$, hence:

$$\frac{a}{b}(0) = \frac{b}{a}(0) \quad \Longrightarrow \quad \frac{a}{b}(0) = \frac{b}{a}(0) = 1$$

This is in close agreement with the observation that the quotient of the off-diagonal terms has an attractor $(0, 1)$ in the function for mesh refinement.

It could be interesting to study the quotient of the off-diagonal coefficients, for the limiting case of an immensely refined grid. It is seen in the above figure that the refinement function becomes more and more dense in the neighbourhood of its attractor. It is suspected that the function may be continuous, or even differentiable, in this region. Assume that for a certain mesh-spacing K there exists a certain number L such that:

$$\frac{b}{a}(K) = L$$

On ground of the theory as it has been developed so far, one can safely write:

$$\frac{b}{a}(K) = \frac{b}{a}\left(2^N.\frac{K}{2^N}\right) = \left[\frac{b}{a}\left(\frac{K}{2^N}\right)\right]^{2^N} = L \quad \Longleftrightarrow \quad \frac{b}{a}\left(\frac{K}{2^N}\right) = L^{1/2^N}$$

By definition, the fraction b/a is continuous for its argument $= 0$ if there does exist a number $\delta(\epsilon)$ in such a way that, for any positive ϵ:

$$\left|\frac{b}{a}(\delta) - \frac{b}{a}(0)\right| < \epsilon$$

But we know that: $b/a(0) = 1$. Identify $\delta = K/2^N$, then the condition for continuity will be fulfilled if:

$$\left|\frac{b}{a}(\delta) - 1\right| = \left|L^{1/2^N} - 1\right| < \epsilon \qquad \Longleftrightarrow$$

$$\left\{\begin{array}{lll} L^{1/2^N} - 1 < \epsilon & \Longleftrightarrow & L < (1+\epsilon)^{2^N} \quad \text{if} \quad L \geq 1 \\ 1 - L^{1/2^N} < \epsilon & \Longleftrightarrow & L > (1-\epsilon)^{2^N} \quad \text{if} \quad L \leq 1 \end{array}\right.$$

$$\left.\begin{array}{lll} \Longleftrightarrow & ln(L) < 2^N.ln(1+\epsilon) & \text{(positive)} \\ \Longleftrightarrow & ln(L) > 2^N.ln(1-\epsilon) & \text{(negative)} \end{array}\right\} \quad \Longleftrightarrow \quad 2^N > \frac{ln(L)}{ln(1 \pm \epsilon)}$$

Where the $\pm$ sign is such that the quotient of the two logarithms is always positive (: $1+\epsilon$ for $L > 1$, $1-\epsilon$ for $L < 1$). We see that, with a suitable choice of $N = N(\epsilon)$:

$$\delta(\epsilon) = \frac{K}{2^N} < K\frac{ln(1 \pm \epsilon)}{ln(L)} \quad \Longrightarrow \quad \left|\frac{b}{a}(\delta) - \frac{b}{a}(0)\right| < \epsilon$$

Which proves that the refinement function for the quotient of the off-diagonal elements is continuous in its attractor.
Now assume that the number ϵ has been selected in such a way that:

$$2^N = \frac{ln(L)}{ln(1 \pm \epsilon)} \quad \Longleftrightarrow \quad 1 \pm \epsilon = L^{1/2^N} \quad \Longleftrightarrow \quad \left\{\begin{array}{ll} \epsilon = L^{1/2^N} - 1 & \text{if} \quad L \geq 1 \\ \epsilon = 1 - L^{1/2^N} & \text{if} \quad L \leq 1 \end{array}\right.$$

Then we find two expressions for the derivative of the Quotient Function in its attractor $(0,1)$:

$$\delta = K\frac{ln(1+\epsilon)}{ln(L)} \quad \text{and} \quad \frac{b}{a}(\delta) - 1 = \epsilon \quad \Longrightarrow \quad \left[\frac{b}{a}(\delta) - \frac{b}{a}(0)\right]/\delta = \frac{\epsilon.ln(L)}{ln(1+\epsilon).K}$$

$$\delta = K\frac{ln(1-\epsilon)}{ln(L)} \quad \text{and} \quad 1-\frac{b}{a}(\delta) = \epsilon \quad \Longrightarrow \quad \left[\frac{b}{a}(\delta) - \frac{b}{a}(0)\right]/\delta = -\frac{\epsilon.ln(L)}{ln(1-\epsilon).K}$$

By continuously enlarging N, we find that $\epsilon \to 0$ and:

$$\lim_{\epsilon\to 0}\frac{\pm\epsilon.ln(L)}{ln(1\pm\epsilon).K} = \frac{ln(L)/K}{\left[ln(1\pm\epsilon) - ln(1)\right]/(\pm\epsilon)} = \frac{ln(L)/K}{ln'(t)|_{t=1}} = \frac{ln(L)}{K}$$

This proves that the Quotient Function has an unique derivative in $(0,1)$. Now it should also be possible to devise a Taylor expansion around the attractor:

$$\frac{b}{a}\left(\frac{dx}{2^N}\right) = \frac{b}{a}(0) + \left[\frac{b}{a}\right]'(0)\,\frac{dx}{2^N} + \ ... = 1 + \left[\frac{b}{a}\right]'(0)\,\frac{dx}{2^N} + \ ...$$

Herewith we can write:

$$\left[\frac{b}{a}\left(\frac{dx}{2^N}\right)\right]^{2^N} \approx \left(1 + \left[\frac{b}{a}\right]'(0)\,\frac{dx}{2^N}\right)^{2^N}$$

A relationship which will become better and better, as the grid becomes more and more refined. In the limiting case, of a "continuous" mesh, we find:

$$\frac{b}{a}(dx) = \lim_{N\to\infty}\left[1 + \left[\frac{b}{a}\right]'(0)\,\frac{dx}{2^N}\right]^{2^N} = e^{\,[b/a]'(0)\,dx}$$

As far a we can see, no special meaning should be attached to the term $[b/a]'(0)$. We found that it is equal to:

$$\left[\frac{b}{a}\right]'(0) = \frac{ln(L)}{K}$$

But L as well as K can be anything, since we assumed that, for a certain mesh spacing K, there exists a starting point L such that $b/a(K) = L$. If we simply put the factor equal to some arbitrary constant $P := ln(L)/K$, then:

$$\frac{b}{a}(dx) = e^{P.dx} = L^{dx/K} \quad \text{and} \quad \frac{a}{b}(dx) = e^{-P.dx}$$

And it is easily shown that known persistence properties for the off-diagonal quotients remain completely unaffected:

$$\left(\left[\frac{b}{a}\right](dx)\right)^2 = \left(e^{P.dx}\right)^2 = e^{P.(2.dx)} = \left[\frac{b}{a}\right](2.dx)$$

$$\left(\left[\frac{b}{a}\right](dx)\right)^{1/2} = \left(e^{P.dx}\right)^{1/2} = e^{P.(dx/2)} = \left[\frac{b}{a}\right](dx/2)$$

Now, indeed, we find the desired generalization of these properties:

$$\frac{b}{a}\,(p.dx) = e^{P.p.dx} = \left[e^{P.dx}\right]^p = \left[\frac{b}{a}\,(dx)\right]^p \quad \text{for any real } p$$

Evidence once more

In a previous section, called 'Quotient Function', evidence has been gathered for the following theorem:

$$\frac{b}{a}(dx) = e^{\,[b/a]'(0)\;dx}$$

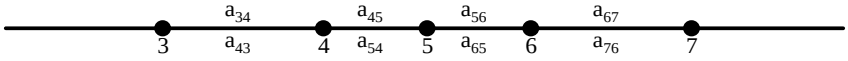
I feel not entirely comfortable with the "proof" in this section, though. The task could be re-formulated as follows: given $f(2.x) = f(x).f(x)$, continuous in x, prove the theorem that: $f(x) = exp(Px)$. It is clear that the theorem is a sufficient condition for $f(2.x) = f(x).f(x)$ being true. But is it also a necessary condition? Up to now, the proof has been accomplished by - sort of - transfinite induction. Truth has been established for grid spacings approaching zero - the domain of Calculus - and then, while working backwards, for finite sized grid spacings, like in Numerical Analysis. In terms of Chaos Theory: the theorem is proved for the Analytical Attractor in a/b(0) = 1 and then very much extrapolated. Therefore trying to arrive at the same result via some other road seems to be worthwile the effort.
The end-result from the section 'Direct Solver' is recalled in the first place:

$$\begin{bmatrix} \cdot & \cdot & \cdot & & & & & \\ & & \cdot & & \cdot & & \cdot & \\ & -\frac{a_{54}}{a_{55}}\frac{a_{43}}{a_{44}} & 0 & 1-\frac{a_{54}}{a_{55}}\frac{a_{45}}{a_{44}}-\frac{a_{56}}{a_{55}}\frac{a_{65}}{a_{66}} & 0 & -\frac{a_{56}}{a_{55}}\frac{a_{67}}{a_{66}} & \\ & & & \cdot & \cdot & \cdot & \\ & & & & \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot \\ \cdot \\ T_3 \\ T_4 \\ T_5 \\ T_6 \\ T_7 \\ \cdot \end{bmatrix}$$

Thus we see that the off-diagonal coefficients of the coarsened grid are given by:

$$a'_{53} = -\frac{a_{54}}{a_{55}}\frac{a_{43}}{a_{44}} \quad \text{and} \quad a'_{57} = -\frac{a_{56}}{a_{55}}\frac{a_{67}}{a_{66}}$$

Independent of the kind of mesh involved. Now imagine a piece of an 1-D mesh again, but this time with a *non-uniform* grid spacing:



Where we have arranged it in such a way that the distance between the vertices (3) and (4) is equal to the distance between the vertices (6) and (7); and the distance between the vertices (4) and (5) is equal to the distance between the vertices (5) and (6). Furthermore it is assumed that the matrix coefficients are only dependent upon these two distances, called:

$$\text{distance}_{\,3-4} = \text{distance}_{\,6-7} = dx_1$$
$$\text{distance}_{\,4-5} = \text{distance}_{\,5-6} = dx_2$$

If matrix elements which are at the left of the main diagonal are denoted as a and matrix elements which are at the right of the main diagonal as b, as has been done before, then it becomes readily evident that:

$$a_{43} = a(dx_1) \quad \text{and} \quad a_{54} = a(dx_2)$$
$$a_{67} = b(dx_1) \quad \text{and} \quad a_{56} = b(dx_2)$$

Furthermore it follows that $a_{44} = a_{66}$ and:

$$a'_{53} = a(dx_1 + dx_2) \quad \text{and} \quad a'_{57} = b(dx_1 + dx_2)$$

Repeat:

$$a'_{57} = -\frac{a_{56}}{a_{55}}\frac{a_{67}}{a_{66}} \quad \text{and} \quad a'_{53} = -\frac{a_{54}}{a_{55}}\frac{a_{43}}{a_{44}}$$

Hence the quotient of the off-diagonal coefficients of the coarsened grid can also be written as follows:

$$\frac{b(dx_1 + dx_2)}{a(dx_1 + dx_2)} = \frac{b(dx_1)}{a(dx_1)}\frac{b(dx_2)}{a(dx_2)}$$

Herewith it is established that the quotient function is characterized by:

$$\frac{b}{a}(dx_1 + dx_2) = \frac{b}{a}(dx_1)\frac{b}{a}(dx_2)$$

And we are triggered to pay some attention to functions which are characterized, in general, by the following property:

$$f(p+q) = f(p).f(q)$$

A couple of other properties can be derived herefrom:

$$f(q) = f(q-p+p) = f(q-p).f(p) \quad \Longrightarrow \quad f(q-p) = f(q)/f(p)$$
$$\Longrightarrow \quad f(p-p) = f(p)/f(p) \quad \Longrightarrow \quad f(0) = 1$$

Employing the hypothesis that f is a differentiable function and taking the limit for $dx \to 0$, we see that:

$$f'(x) = \frac{f(x+dx) - f(x)}{dx} = \frac{f(x).f(dx) - f(dx)}{dx} =$$

$$f(x)\frac{f(dx) - 1}{dx} = f(x)\frac{f(dx) - f(0)}{dx} = f'(0).f(x)$$

Giving a differential equation for $f(x)$ which can be solved rather easily:

$$\frac{df}{f} = f'(0).dx \quad \Longrightarrow \quad ln(f) = f'(0).x + c \quad \Longrightarrow \quad f(x) = Ce^{f'(0).x} = e^{f'(0).x}$$

Where the boundary condition $f(0) = 1$ is employed in the last step. $f'(0) = P$ is an arbitrary constant. The result is in agreement with our conjecture that the only possible Quotient Functions are, indeed, given by: $a/b(dx) = exp(P.dx)$. If the above proof is valid, then this conjecture has been turned into a theorem.

Some Stable Solutions

A Finite Difference scheme is associated with any (normed) tri-diagonal system of equations in the following manner:

$$-b.T_{i-1} + T_i - a.T_{i+1} = 0$$

The equations, when put in this form, can be solved in a direct fashion. Let's try the most simple of all possible solutions: a constant. Substitution of $T_i = C$ gives:

$$-b.C + C - a.C = 0 \quad \Longrightarrow \quad a + b = 1$$

Assuming that $C \neq 0$ (which would result in an even more trivial solution). It is seen that a constant solution can be obtained only if (minus) the sum of the off-diagonal elements is one.
The next trivial solution of the finite difference equation would be a straight line. Substitute $T_i = C.i + D$ with $D \neq 0$ and $C \neq 0$, giving:

$$-b.\,[C.(i-1) + D] + [C.i + D] - a.\,[C.(i+1) + D] =$$

$$C.i(-b + 1 - a) + D(-b + 1 - a) + (b - a) = 0$$

Resulting in the same condition $a + b = 1$, augmented with $b - a = 0$. It is concluded that a *linear solution* can only be obtained for:

$$a = b = \frac{1}{2}$$

Let another possible solution be given as:

$$T_i = \left(\frac{b}{a}\right)^i$$

And substitute. Then:

$$-b\left(\frac{b}{a}\right)^{i-1}+\left(\frac{b}{a}\right)^{i}-a\left(\frac{b}{a}\right)^{i+1}=0 \quad \Longrightarrow$$

$$(b/a)^i(-b.a/b+1-a.b/a)=(b/a)^i(-b+1-a)=0$$

Resulting again in the condition $a+b=1$. A more general solution may be cast into the form:

$$T_i=C\left(\frac{b}{a}\right)^{i}+D$$

Suppose that we have boundary conditions, like:

$$T_1=1 \quad \text{and} \quad T_N=0$$

Then the constants C and D can be calculated explicitly, resulting in:

$$T_i=\frac{(b/a)^i-(b/a)^N}{(b/a)-(b/a)^N}=\frac{(b/a)^{i-1}-(b/a)^{N-1}}{1-(b/a)^{N-1}}$$

This result is more general than it seems at first sight, because it may be multiplied with an arbitrary constant and another arbitrary constant may be added to it, thus adapting it to quite arbitrary boundary conditions:

$$T_i:=(T_1-T_N).T_i+T_N$$

It is recognized that the condition $a+b=1$ plays a rather predominant role in its relationship to the solutions which are found so far. The meaning of the condition is that the sum of the row coefficients in the (normed) matrix equals zero. Things are clarified further by writing the equations in a slightly different manner:

$$T_i=b.T_{i-1}+a.T_{i+1}$$

Knowing that $b\geq 0$, $a\geq 0$ and $a+b=1$. It is concluded therefrom that T_i is a *weighted mean* of its neighbouring values, which means that it will always lie in between T_{i-1} and T_{i+1}:

$$T_{i-1}\leq T_i\leq T_{i+1} \quad \text{or} \quad T_{i+1}\leq T_i\leq T_{i-1}$$

Hence, irrespective of any boundary conditions, the function behaviour of the numerical solution will be continuously decreasing or continuously increasing. It shows no "wiggles". Such a solution T_i is commonly called *stable*. Thus, in fact, we are in search here for Stable Solutions of the tri-diagonal equations. It is remarked, in addition, that stability conditions, like $a+b=1$ for the 1-D case, are well known and commonly accepted, as they belong to the "Four Basic

Rules" in Patankar's book (1980).
Having established stability, our Stable numerical Solution is recalled:

$$T_i = \frac{(b/a)^{i-1} - (b/a)^{N-1}}{1 - (b/a)^{N-1}}$$

Because, with help of the "Quotient Function", it can be cast now into the following form:

$$T_i = \frac{e^{P.(i-1).dx} - e^{P.(N-1).dx}}{1 - e^{P.(N-1).dx}} \quad \Longleftrightarrow \quad T(x) = \frac{e^{P.x} - e^{P.L}}{1 - e^{P.L}}$$

Where $x = (i-1).dx$ is the 1-D coordinate and L is the total length of the mesh. Herewith, the general Stable (: $a + b = 1$) numerical Solution becomes quite alike an Analytical one:

$$T(x) = (T_0 - T_L)\frac{e^{P.x} - e^{P.L}}{1 - e^{P.L}} + T_L$$

This solution is applicable to 1-D problems of Convection and Diffusion. It is concluded therefrom that any tri-diagonal system of linear equations, where the off-diagonal elements are negative and the rows sum up to zero, is actually a blueprint for Convection and Diffusion.
The number $P.L$ in the exponents is recognized as the *Péchlet number*. In real life, the dimensionless Péchlet number has the following meaning: $Pe = \rho.c.v.L/\lambda$, where $\rho =$ density (kg/m^3), $c =$ heat capacity $(J/kg/K)$, $v =$ velocity (m/s) and $\lambda =$ thermal conductivity $(J/s/m/K)$.
For the problem of 1-D Convection and Diffusion, the off-diagonal coefficients a and b can be written in a form which will be useful in the subsequent work. Define ϵ by:

$$\epsilon = b - a \qquad \text{while knowing} \qquad a + b = 1$$

By addition and subtraction we find:

$$b = \frac{1}{2}(1 + \epsilon) \quad \text{and} \quad a = \frac{1}{2}(1 - \epsilon) \quad \text{where} \quad -1 \leq \epsilon \leq +1$$

It is remarked that the range of ϵ is quite restricted in the first place. Furthermore it can be argued that, as the grid spacing becomes smaller and smaller with successive refinements of the grid, then ϵ also must become smaller and smaller:

$$\frac{1 - \epsilon}{1 + \epsilon} = \frac{a}{b}(dx \to 0) \to 1$$

With help of a clever trick, we can express ϵ solely in quotients of the off-diagonal coefficients. Divide namely nominator and denominator by $\sqrt{a.b}$ in:

$$\epsilon = \frac{b - a}{b + a} = \frac{\sqrt{b/a} - \sqrt{a/b}}{\sqrt{b/a} + \sqrt{a/b}} = \frac{e^{+P/2.dx} - e^{-P/2.dx}}{e^{+P/2.dx} + e^{-P/2.dx}}$$

The latter expression is recognized as a Hyperbolic Tangent: $tanh(P/2.dx)$, as will be seen in one of the subsequent paragraphs. A sketch of the $tanh$ function reveals that ϵ will have the following properties:

$$\lim_{P\to-\infty} \epsilon(P) = -1 \qquad \epsilon(0) = 0 \qquad \lim_{P\to+\infty} \epsilon(P) = +1$$

When cast in Finite Element form, the analytical Convection-Diffusion scheme reads as follows:

$$\begin{bmatrix} +a & -a \\ -b & +b \end{bmatrix} = \begin{bmatrix} +\frac{1}{2} - \frac{1}{2}\epsilon & -\frac{1}{2} + \frac{1}{2}\epsilon \\ -\frac{1}{2} - \frac{1}{2}\epsilon & +\frac{1}{2} + \frac{1}{2}\epsilon \end{bmatrix}$$

The trace of any such element matrix is 1. Double check for pure diffusion and for pure convection, respectively:

$$P = 0 \quad \Longrightarrow \quad \begin{bmatrix} +a & -a \\ -b & +b \end{bmatrix} = \begin{bmatrix} +1/2 & -1/2 \\ -1/2 & +1/2 \end{bmatrix}$$

$$P = \infty \quad \Longrightarrow \quad \begin{bmatrix} +a & -a \\ -b & +b \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -1 & +1 \end{bmatrix}$$

$$P = -\infty \quad \Longrightarrow \quad \begin{bmatrix} +a & -a \\ -b & +b \end{bmatrix} = \begin{bmatrix} +1 & -1 \\ 0 & 0 \end{bmatrix}$$

The latter expressions are demanding for a boundary condition at the inlets $x = 0$ and $x = L$ respectively. Otherwise, the temperature $T(x)$ at these places would be undefined.

Product Function

The *product* of the off-diagonal coefficients at the coarsened grid can be expressed as a function of the product of the off-diagonal terms at the refined grid:

$$a'.b' = \frac{a^2}{1-2.a.b}\,\frac{b^2}{1-2.a.b} = \left(\frac{a.b}{1-2.a.b}\right)^2$$

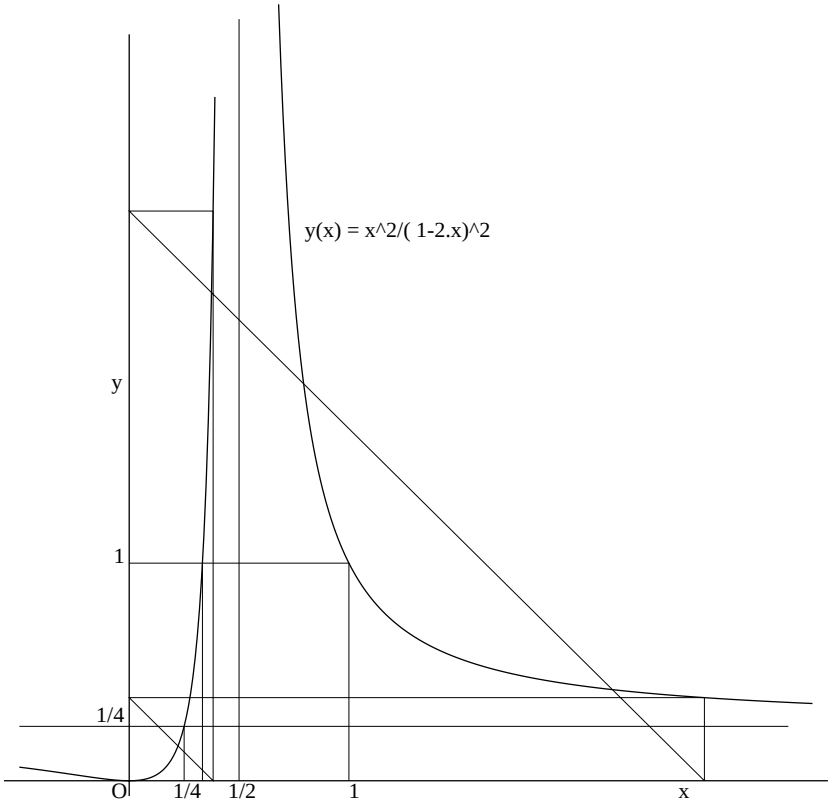
Or:

$$y = \left(\frac{x}{1-2.x}\right)^2 \qquad \text{where} \qquad y = a'.b' \quad \text{and} \quad x = a.b$$

At first, we make a sketch of the function which relates the product $a'.b'$ on the finer grid to the product $a.b$ on the coarser grid. That is, the function $y = x^2/(1-2.x)^2$. The function has two asymptotes, a vertical one for $x = 1/2$ (denominator $= 0$) and a horizontal one for $y = 1/4$:

$$\lim_{x\to\pm\infty} \left(\frac{x}{1-2.x}\right)^2 = \lim_{x\to\pm\infty} \left(\frac{1}{2-1/x}\right)^2 = \frac{1}{4}$$

We want to make a small step aside now, and ask attention for a subject which is on the experimental side and therefore a bit different from the mainstream theoretical argument. While having a look at the graphics, the reader may wonder how these pictures were created. The answer is that most of the figures were composed by making use of *native PostScript*. This seems to be weird at first sight, but it should be remarked here that PostScript is, in fact, a full blown programming language. (Moreover, it is a language which offers no special difficulties to someone who has been an experienced FORTH programmer.) Much of the theory in this document has been accompanied by numerical experiments. These experiments were implemented and carried out a great deal in (Turbo) Pascal. But, since PostScript is a programming language, nothing has prevented us from having carried out some of these numerical experiments entirely in PostScript. That's why the source code of the pictures in this document actually contain more information than may be displayed. It is suggested herewith that the PSF (PostScriptFile) programs, as delivered with this document, may be interesting as such. Having said this, let's return to theory, starting with the figure below.



With grid coarsening, the function is used *iteratively*. Get started with a certain

value of x and the accompanying value of y can be calculated. But this value of y serves as the next x value for the function y, as soon as the grid is coarsened again. In general, we obtain an assembly of the form:

$$y\,(y\,(y\,(y\,(y...y(x)))))$$

It may be questioned if there exist values of x which are are such that, after coarsening of the grid, the same value $x = y(x)$ is acquired again. Such values will remain the same during all stages of the grid coarsening process. For this reason, such values of x are called *stationary* or *invariant* points of the function. They are found as follows:

$$x = \left(\frac{x}{2.x-1}\right)^2 \quad \Longrightarrow \quad x = 0 \quad \text{or} \quad 1.(2.x-1)^2 = x \quad \Longrightarrow$$

$$x^2 - \frac{5}{4}x + 1 = (x-1)(x-\frac{1}{4}) = 0$$

Herewith invariant points are found to be:

$$x = 0 \quad \text{or} \quad x = 1 \quad \text{or} \quad x = \frac{1}{4}$$

The value $1/3$, being useful because of the condition $a.b < 1/3 < 1/2$, is mapped on the stationary value 1:

$$\left(\frac{1/3}{1-2.1/3}\right)^2 = 1$$

But the above list of invariant points is not exhaustive, another possibility being that there exist *two* points instead of one. After having obtained the point with label (1) in one iteration, the point with label (2) is reached in the next iteration. Then the point with label (1) is obtained again, and so on and so forth. We must solve an equation with $1/x$ instead of x:

$$\frac{1}{x} = \left(\frac{x}{2.x-1}\right)^2 \quad \Longrightarrow \quad (2.x-1)^2 = x^3 \quad \Longrightarrow \quad x^3 - 4.x^2 + 4.x - 1 = 0$$

An equation of the third degree. However, we already know that $x = 1/x = 1$ must be a solution of it. Hence we can perform a division:

$$\frac{x^3 - 4.x^2 + 4.x - 1}{x-1} = x^2 - 3.x + 1 = 0$$

The roots of the equation are herewith found to be:

$$x_1 = 1 \quad \text{or} \quad x_2 = \frac{3}{2} + \frac{\sqrt{5}}{2} \approx 2.618 \quad \text{or} \quad x_3 = \frac{3}{2} - \frac{\sqrt{5}}{2} \approx 0.382$$

Indeed: $x_1 = 1/x_1$, $x_2 = 1/x_3$ and $x_3 = 1/x_2$. See the above figure.
Another interesting property of the invariant point $(1/4, 1/4)$ can be found as follows. Consider the the expression $(1 - 4.a.b)$. We will derive a persistent property for it, by considering the expression while refining the mesh. We have found, in the previous chapter, that the product of the off-diagonal matrix coefficients is transformed as follows:

$$a.b = \frac{\sqrt{a'.b'}}{1 + 2.\sqrt{a'.b'}}$$

Herewith we find that $1 - 4.a.b$ is also transformed, according to:

$$1 - 4.a.b = \frac{1 + 2.\sqrt{a'.b'} - 4.\sqrt{a'.b'}}{1 + 2.\sqrt{a'.b'}} = \frac{\left(1 - 2.\sqrt{a'.b'}\right)\left(1 + 2.\sqrt{a'.b'}\right)}{\left(1 + 2.\sqrt{a'.b'}\right)^2} =$$

$$\frac{1 - 4.a'.b'}{\left(1 + 2.\sqrt{a'.b'}\right)^2}$$

Meaning that the *sign* of the $1 - 4.a.b$ is insensitive/invariant for for mesh refinement, persistent through multigrids:

$$\begin{aligned} 1 - 4.a.b \geq 0 \quad &\Longleftrightarrow \quad 1 - 4.a'.b' \geq 0 \\ 1 - 4.a.b \leq 0 \quad &\Longleftrightarrow \quad 1 - 4.a'.b' \leq 0 \end{aligned}$$

Thus if the expression is positive, then it remains positive. And if it is negative, then it remains negative. Last but not least, if it is zero, then it remains zero, the latter corresponding with the invariance of $(1/4, 1/4)$. This may be translated as follows:

$$a.b \leq \frac{1}{4} \quad \Longleftrightarrow \quad a'.b' \leq \frac{1}{4} \qquad \text{and} \qquad a.b \geq \frac{1}{4} \quad \Longleftrightarrow \quad a'.b' \geq \frac{1}{4}$$

Hence the point $(1/4, 1/4)$ serves as a boundary between two domains: points in the domain $x > 1/4$ will always give rise to other points x with $x > 1/4$, while points in the domain $x < 1/4$ will always give rise to other points x with $x < 1/4$. That's why **discriminant** is probably a good name for the expression $(1 - 4.a.b)$. With the sign of the discriminant, one can discriminate between two seemingly distinct domains of interest in $(a.b)$-space.
The properties $x' < x$ or $x' > x$ are persistent at MultiGrids. Let's prove it:

$$x' = x^2/(1-2.x)^2 < x \quad \Longleftrightarrow \quad 1-4.x+4.x^2 > x \quad \Longleftrightarrow \quad x^2-5/4.x+1/4 > 0$$

The accompanying function is a parabola:

$$y = \left(x - \frac{5}{8}\right)^2 - \frac{25}{64} + \frac{16}{64}$$

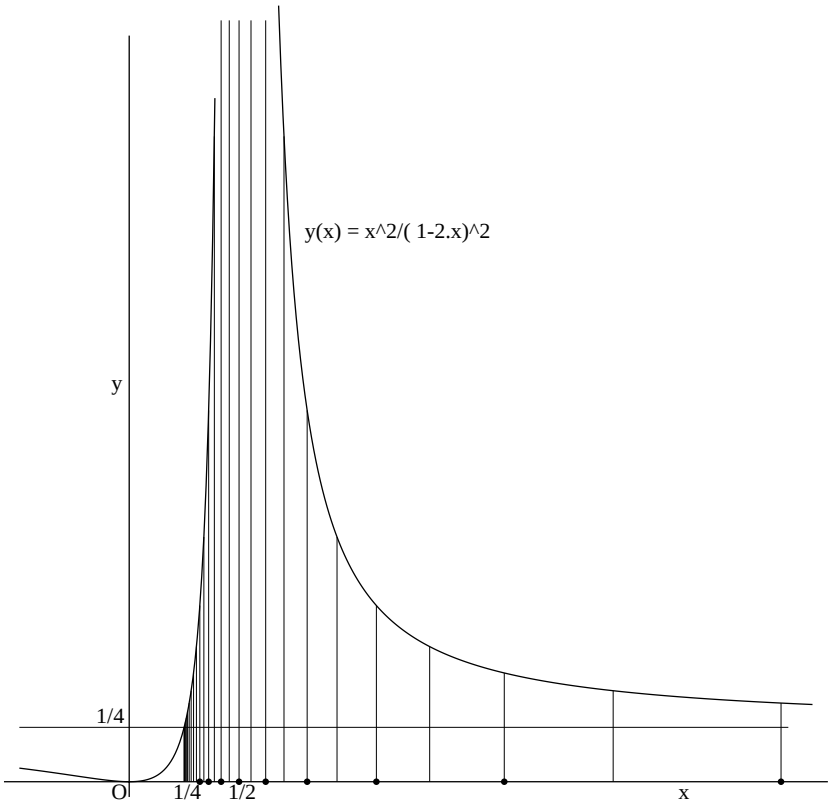
Which has a minimum for $(x, y) = (5/8, -9/64)$. It is positive for:

$$(x-1)(x-\frac{1}{4}) > 0$$

Ignoring the conditions $x > 1$ and $x < 1$, while $x < 1/3 < 1/2$, results in:

$$x > x' \iff x < \frac{1}{4} \quad ; \quad x < x' \iff x > \frac{1}{4}$$

This means that products smaller than $1/4$ will become bigger and products bigger than $1/4$ will become smaller with grid refinement; any product of coefficients will approximate $1/4$ closer with successive refinement at uniform Multi-Grids. We have seen that the latter outcome is persistent through further grid refinement. With other words, the point $1/4$ is a stationary point (attractor) of the Product Function for grid refinement. It can also be said that $1/4$ *repels* the products during the reverse process, of grid coarsening.



During a coarsening process, points x may give rise to other points $x = y(x)$ which are such that x approaches the value $1/2$, in the end. For this value the denominator $(1 - 2.x)$ becomes zero. Points x such as mentioned will be called **dangerous** in the sequel. Let's forget for a moment the somewhat premature discovery that $x < 1/3 < 1/2$. And let the search for dangerous points start here from scratch:

$$x' = \left(\frac{x}{1-2.x}\right)^2 \quad \Longrightarrow \quad \sqrt{x'} = \frac{x}{1-2.x} = \frac{x}{2.x-1} \quad \Longrightarrow$$

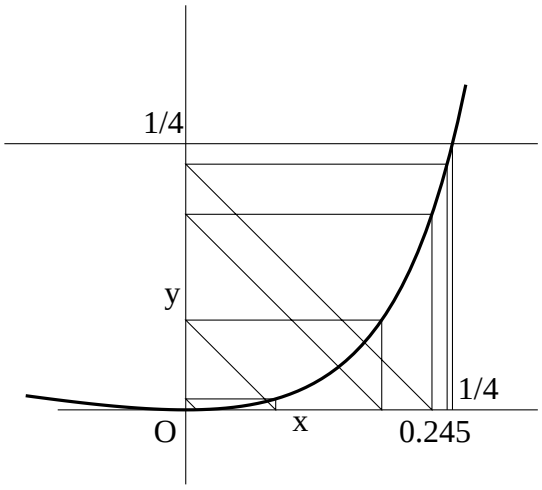
$$\pm\, x = \sqrt{x'}\,(2.x-1) \quad \Longrightarrow \quad x\,(2.\sqrt{x'} \pm 1) = \sqrt{x'} \quad \Longrightarrow$$

$$x = \frac{\sqrt{x'}}{2.\sqrt{x'} \pm 1} = \frac{1}{2 \pm 1/\sqrt{x'}}$$

Hence each dangerous point will give rise to 2 other dangerous points. After 4 iterations, for example, there will be $1 + 2 + 4 + 8 + 16 = 31$ dangerous points. It seems useful to know how these points are distributed along the x-axis. Some experimental results are depicted in the above figure.
It is conjectured that the whole area $1/4 < x < +\infty$ is, in fact, more or less dangerous, with exception perhaps of a few stationary points. But even for such invariant points, there is always the danger of becoming less stationary, because of roundoff errors. The latter will inevitably be present, as soon as numbers are calculated in a real world environment.
If the above is true, then the only "safe" domain for x while iterating with the function $y(x)$ would be given by $0 \leq x \leq 1/4$. Better substitute the values $x = a.b$ and $y = a'.b'$. Then we find, indeed, that the denominator will never be zero:

$$0 \leq a.b \leq \frac{1}{4} \quad \Longrightarrow \quad \frac{1}{2} \leq 1 - 2.a.b \leq 1$$

And we know for sure that $a.b$ will never "escape" from the interval $(0, 1/4)$.
If a starting value of $x = a.b$ is selected somewhere inside the interval $0 \leq x < 1/4$, and it does *not* coincide with the point $1/4$, then it is observed that the successive iterates converge (rather quickly) to $x = 0$. If we commit ourselves to modern terminology, then we would say that the point $x = 0$ is an *attractor* of the function $y(x) = x^2/(2.x-1)^2$ over the interval $0 \leq x < 1/4$. An enlargement of the function in the neighbourhood of the attractor is depicted in the figure on the next page.



"Some Stable Solutions" were found in the chapter with the same name, but they were found regardless of any considerations about a "safe" domain of interest. Is it a coincidence that the safety condition $a.b \leq 1/4$ does not play any role in this case? Is it perhaps due to the fact that the stability condition we have adopted instead, is actually stronger than the condition for safety? Let's see:

$$a + b \leq 1 \quad \Longrightarrow \quad (a+b)^2 \leq 1 \quad \Longrightarrow \quad a^2 - 2.a.b + b^2 \leq 1 - 4.a.b$$
$$\Longrightarrow \quad 1 - 4.a.b \geq (a-b)^2 \geq 0$$

Herewith it is demonstrated that Stability is, indeed, a sufficient condition for Safety. But it should be emphasized, at the same time, that Stability is *not necessary* for Safety: there may exist Safe Solutions which are not Stable.

The Trigonometric Connection I

A bit of elementary trigonometry is needed in order to acquire further knowledge. Start with:

$$cos(2.\phi) = 2.cos^2(\phi) - 1$$

Solve for $cos(\phi)$, divide ϕ by two and take care of the signs:

$$cos(\frac{1}{2}\phi) = \sqrt{\frac{1}{2} + \frac{1}{2}cos(\phi)} \quad \text{for} \quad 0 \leq \phi \leq \pi$$

Augmented with:

$$cos(\pi - \phi) = -cos(\phi)$$

But suppose we are rather interested in the function $D(h) = 2 + 2.cos(\pi.h)$, restricted to the range $0 \leq h \leq 1$. Rewrite the above formula as such:

$$2 + 2.cos(\pi.\frac{1}{2}h) = 2 + \sqrt{2 + 2.cos(\pi.h)} \quad \text{for} \quad 0 \leq h \leq 1$$

It follows that:

$$D(\frac{1}{2}h) = 2 + \sqrt{D(h)} \quad \text{for} \quad 0 \leq h \leq 1$$

Augmented with:

$$2 + 2.cos\,[\pi(1-h)] = 2 - 2.cos(\pi.h) \qquad \text{or} \qquad D(1-h) = 4 - D(h)$$

The latter formula can also be implemented in a more "symmetric" way, as opposed to $D(h/2) = 2 + \sqrt{D(h)}$:

$$D(1 - \frac{1}{2}h) = 4 - D(\frac{1}{2}h) = 2 - \sqrt{D(h)} \quad \text{for} \quad 0 \leq h \leq 1$$

Elementary values are:

$$D(0) = 4 \qquad D(\frac{1}{2}) = 2 \qquad D(1) = 0$$

Start with $h = 1/2$ then:

$$D(\frac{1}{4}) = 2 + \sqrt{D(\frac{1}{2})} = 2 + \sqrt{2} \qquad D(\frac{3}{4}) = 4 - D(\frac{1}{4}) = 2 - \sqrt{2}$$

Let's try now for fractions $\times 1/8$. The values D(0/8), $D(2/8)$, D(4/8), D(6/8) and D(8/8) have already been calculated. Go for the rest:

$$D(\frac{1}{8}) = 2 + \sqrt{D(\frac{1}{4})} = 2 + \sqrt{2 + \sqrt{2}} \qquad D(\frac{7}{8}) = 4 - \sqrt{D(\frac{1}{8}} = 2 - \sqrt{2 + \sqrt{2}}$$

$$D(\frac{3}{8}) = 2 + \sqrt{D(\frac{3}{4})} = 2 + \sqrt{2 - \sqrt{2}} \qquad D(\frac{5}{8}) = 4 - D(\frac{3}{8}) = 2 - \sqrt{2 - \sqrt{2}}$$

The outcomes can be *sorted*, in *descending* order, or with the h-values as a key:

$$\begin{aligned}
D(0/8) &= 4 \\
D(1/8) &= 2 + \sqrt{2 + \sqrt{2}} \\
D(2/8) &= 2 + \sqrt{2} \\
D(3/8) &= 2 + \sqrt{2 - \sqrt{2}} \\
D(4/8) &= 2 \\
D(5/8) &= 2 - \sqrt{2 - \sqrt{2}} \\
D(6/8) &= 2 - \sqrt{2} \\
D(7/8) &= 2 - \sqrt{2 + \sqrt{2}} \\
D(8/8) &= 0
\end{aligned}$$

It is evident that such a procedure will work for any denominator of the form 2^N where $N > 0$ is an integer. The proof is by complete induction. Suppose that the work has already been done for 2^{N-1}, then all function values for angles $\pi.2.k/2^N$ are known. The values for $\pi.k/2^N$ can be calculated by using $D(k/2^N) = 2 + \sqrt{D(2.k/2^N)}$, giving all values $k/2^N$ in the interval $0 < k/2^N < 1/2$. The remaining values, in the interval $1/2 < k/2^N < 1$, can then be calculated by using $D(1 - k/2^N) = 4 - D(k/2^N)$.
The structures of square-roots-of-two are very much alike, except for the $\pm$ signs. Hence they may be abbreviated as follows and interpreted as binary codes or "numbers":

$$2 + \sqrt{2 - \sqrt{2 + \sqrt{2 + \sqrt{2 - ...}}}} \quad \equiv \quad + - + + - ... \quad \equiv \quad 0\,1\,0\,0\,1 ...$$

Calculations for denominators of higher degree can be automated by a computer program. By doing so, it is observed that the "new" numbers, with k = odd, together form a *binary-reflected Gray-code*. A Gray-code is characterized by the property that only *one* "bit" is changed at a time, while going to the next codeword (provided that both codewords are of equal length). Suppose we have a N-bit Gray code which is represented by a $(N \times 2^N)$ matrix of '+'s and '−'s, in such a way that the i'th column is the i'th codeword:

$$G(N) = \left[\begin{array}{cccccc} G_0 & G_1 & ... & G_i & ... & G_{2^N-1} \end{array} \right]$$

The order of the codes belonging to $G(N)$ remains unchanged while using the formula $D(h/2) = 2 + \sqrt{D(h)}$, the order is reversed while using the formula $D(1 - h/2) = 2 - \sqrt{D(h)}$, for obtaining the next generation of codewords. Thus, while performing the next step of the angle-refinement procedure, the code is augmented, for odd indices k, with '+' and '−' signs as follows:

$$G(N+1) = \left[\begin{array}{cccccccccc} G_0 & G_1 & ... & G_{2^N-1} & G_{2^N-1} & G_{2^N-2} & ... & G_1 & G_0 \\ + & + & ... & + & - & - & ... & - & - \end{array} \right]$$

If $G(N)$ is a N-bit Gray code, then herewith it is clear that $G(N+1)$ must be a $(N+1)$-bit Gray code. Our proof by induction is completed by verifying the existence of a trivial 1-bit code:

$$G(1) = \left[\begin{array}{cc} + & - \end{array} \right]$$

Let's consider again now the functions which relate the the product of the off-diagonal coefficients at the coarser grid to the product of coefficients at the finer grid and vice versa. Start with:

$$x = \frac{\sqrt{x'}}{2.\sqrt{x'} \pm 1} = \frac{1}{2 \pm 1/\sqrt{x'}}$$

The accompanying process of grid refinement may also be described as follows:

$$\frac{1}{x} = 2 \pm \sqrt{\frac{1}{x'}}$$

Start with $x' = 1/2$ or $1/x' = 2$, the value for which the (inverse) coarsening process explodes, then $1/x = 2 \pm \sqrt{2}$. Iterating again gives: $1/x := 2 \pm \sqrt{2 \pm \sqrt{2}}$. And so on and so forth. If these values are sorted in descending order, and plotted on a computer screen, then the resemblance with a cosine function readily becomes apparent. Indeed, the sequence:

$$2 \pm \sqrt{2 \pm \sqrt{2 \pm \sqrt{2 \pm \sqrt{2 \pm \sqrt{2...}}}}}$$

is recognized to be quite the same one as has been produced by iterations with the function D: start with $h = 1/2$ and do exactly the same with:

$$D(\frac{1}{2}h) = 2 + \sqrt{D(h)} \quad \text{or} \quad D(1 - \frac{1}{2}h) = 2 - \sqrt{D(h)}$$

We may conclude that there exists an intimate, one-to-one relationship between two seemingly quite different functions. The first one maps the catastrophic value $1/2$ onto points in the space of products of off-diagonal coefficients, which may then called "dangerous". The second one calculates all successive values of $2 + 2.cos(k.\pi/2^N)$ for all k with $0 \leq k \leq 2^N$ and N an arbitrary large integer, starting with the outcomes $2 \pm \sqrt{2}$. In fact, nothing prevents us from writing:

$$\frac{1}{a'.b'} \equiv D(h) \quad \text{and} \quad \frac{1}{a.b} \equiv D(\frac{1}{2}h)$$

Where $a.b$ denotes the off-diagonal coefficients product, which is obtained by refinement of the grid belonging to $a'.b'$. It is observed that grid-refinement corresponds with *halving an angle*. It may be questioned if the reverse is also true: does grid coarsening correspond with *doubling* an angle?

$$x' = \left(\frac{x}{2.x - 1}\right)^2 \quad \Longleftrightarrow \quad \frac{1}{x'} = \left(2 - \frac{1}{x}\right)^2$$

Which is equivalent with:

$$cos(2.\phi) = 2.cos^2(\phi) - 1 \quad \Longleftrightarrow \quad 2 + 2.cos(2.\phi) = 2 + 4.cos^2(\phi) - 2 =$$

$$[2.cos(\phi)]^2 = (2 - [2 + 2.cos(\phi)])^2 \quad \Longleftrightarrow \quad D(2.h) = [2 - D(h)]^2$$

Where $\phi = \pi.h$. Because $1/x \equiv h$ and $1/x' \equiv 2.h$, there is enough solid ground now to identify:

$$\frac{1}{a.b} = D(h) = 2 + 2.cos(\pi.h)$$

Herewith we find the distribution of dangerous points in $(a.b)$-space, belonging to a grid refinement of order N, as has been verified by numerical experiments too:

$$a.b = \frac{1}{2 + 2.cos(k.\pi/2^N)} \qquad k = 0, 1, 2, ..., 2^N$$

The Trigonometric Connection II

A number of conclusions can be drawn by properly employing the formula:

$$a.b = \frac{1}{2 + 2.cos(k.\pi/2^N)} \qquad k = 0, 1.2, ..., 2^N$$

The function reaches a minimum $1/4$ for the maximum value of the denominator, which is 4. Because the minimum of the denominator is zero, it also reaches to infinity, meaning that dangerous points can be found *at arbitrary density*, increasing with N, and *everywhere* in the region $1/4 < a.b < \infty$. It is concluded therefore that a region which is safe everywhere can only exist in the interval $0 \leq a.b \leq 4$, as we have seen (but not proved) before.
However, not all points in the region $1/4 < a.b < \infty$ share the same amount of risk. It seems reasonable to conjecture that points which are very close to $1/4$ are less "dangerous" that others. It will be shown now that this is indeed true. Points close to $1/4$ correspond with small angles. Hence any coarsening of the grid will correspond with the doubling of a small angle. The closer the point is to $1/4$, the smaller the angle will be. And hence it will require quite some doubling effort, before this point comes in real danger. To be more specific. Suppose that $a.b$ almost equals $1/4$. We want to use a Taylor expansion here. Some hand-held calculus can be avoided by devising an input for our favourite Computer Algebra System (MAPLE):

```
series(1/(2+2*cos(x)),x);
```

$$\frac{1}{4} + \frac{1}{16}x^2 + \frac{1}{96}x^4 + O(x^6)$$

In our case, suppose that we have a point $1/4 + \delta$, corresponding with an angle $\pi.1/2^N$. Then:

$$\frac{1}{4} + \delta \approx \frac{1}{4} + \frac{1}{16}(\pi/2^N)^2$$

Solving for the angle:

$$\pi/2^N \approx \sqrt{16.\delta} \quad \Longrightarrow \quad ln(\pi) - N.ln(2) \approx 0.5\, ln(16) + 0.5\, ln(\delta) \quad \Longrightarrow$$

$$N \approx 0.5\, {}^2log(1/\delta) \quad + \quad {}^2log(\pi) - 2$$

If, for example, the deviation from $1/4$ is one part per million $= 10^{-6}$, then $\delta \approx 2^{-20} \rightarrow 0.5\, {}^2log(1/\delta) \approx 10$. This means that a grid coarsening less than

10 times offers no real danger. Which in turn means that the finest mesh can be allowed to consist of $2^{10} \approx 1000$ points, without possibly running into big trouble.
What more can be said about the stationary points we have found in the past:

$$\frac{1}{4} \quad , \quad 1 \quad , \quad \frac{3 \pm \sqrt{5}}{2}$$

Or, solve for the angle ϕ in:

$$a.b = \frac{1}{2 + 2.cos(\phi)} = \quad \frac{1}{4} \quad , \quad 1 \quad , \quad \frac{3}{2} \pm \frac{\sqrt{5}}{2}$$

A useful remark being:

$$\left(\frac{3}{2} + \frac{\sqrt{5}}{2}\right)\left(\frac{3}{2} - \frac{\sqrt{5}}{2}\right) = 1$$

Solve for the cosines first:

$$cos(\phi) = \quad 1 \quad , \quad -\frac{1}{2} \quad , \quad \frac{\pm\sqrt{5} - 1}{4}$$

The first angle is just zero: $\phi = 0$ and any doubling or halving of this angle maps it upon itself. Hence 1/4 is a single stationary point, as we have seen.
The second angle must be 120^0. A doubling of this angle gives $cos(240^0) = -1/2$, which is the same value as with $\phi = 120^0$. Hence 1 is also a single stationary point, as we have seen.
The 3rd/4th angles are somewhat more complicated. The smallest of the two will correspond with the plus sign. Doubling this one gives:

$$cos(2.\phi) = 2.cos^2(\phi) - 1 = 2\left(\frac{+\sqrt{5} - 1}{4}\right)^2 - 1 = \frac{5 - 2\sqrt{5} + 1}{8} - 1 = \frac{-\sqrt{5} - 1}{4}$$

Which is precisely the second one. Let's double it again:

$$cos(4.\phi) = 2.cos^2(2.\phi) - 1 = 2\left(\frac{-\sqrt{5} - 1}{4}\right)^2 - 1 = \frac{5 + 2\sqrt{5} + 1}{8} - 1 = \frac{+\sqrt{5} - 1}{4}$$

Giving back the first of the two. We find that the angles are determined by:

$$cos(4.\phi) = cos(\phi) \quad \Longrightarrow \quad 4.\phi = k.2.\pi \pm \phi \quad \Longrightarrow$$

$$\phi = k\frac{2.\pi}{5} \quad \left(\quad \text{or} \quad \phi = k\frac{2.\pi}{3} \right) \qquad \text{where} \quad k = 1, 2, 4, 8, \ ...$$

The product-function values corresponding with the denominator 5 are stationary with a multiplicity of two: the first point gives the second point, the second point gives back the first point, and so on and so forth.
It is questioned now whether there exist also points which have a multiplicity of three or higher. The answer is affirmative. Take for example an angle which gives the same function value again, but only after three times doubling itself:

$$cos(8.\phi) = cos(\phi) \quad \Longrightarrow \quad 8.\phi = k.2.\pi \pm \phi \quad \Longrightarrow$$

$$\phi = k\frac{2.\pi}{7} \quad \text{or} \quad \phi = k\frac{2.\pi}{9} \qquad \text{where} \quad k = 1, 2, 4, 8, \ ...$$

We have found some clues for the behaviour of points in the area $(a.b) > 1/4$, corresponding with angles $k.2.\pi/D$ and denominators $D = 2, 3, 4, 5, 7, 8, 9, ...$ A denominator like 6 can be produced by $1/3 = 2/6$ and halving the latter. The same story can be told for $10 : 1/5 = 2/10$. Denominators 31 and 33 can be produced by $2^5.\phi = k.2.\pi \pm \phi$, corresponding with multiplicities ≤ 5. Then $3/33 = 1/11$, providing material for angles $k.2.\pi/11$.
It is thus observed that, in the "dangerous" area, there also exist infinitely many points which are not "dangerous" at all, since they will always be mapped upon each other with any further grid coarsening. But, as has been suggested earlier, this situation is similar to *unstable equilibrium* in physics, like a pencil balancing on its tip.
Fractions $k/2^N$ are expanded in a computer as binary numbers. For example:

0.00101001010011101010110101110110

But also numbers like $1/3$ are represented, in a computer, as a finite binary fraction:

0.01010101010101010101010101010101

Given a finite (though maybe very large) binary representation of any number in the area $1/4 < a.b < \infty$, there is no sensible way to tell if this number is predestined to be dangerous or not. This remains true, even if we ever would have a computer at our disposal with immensely large words.
This story vaguely reminds to the (in)famous controversy between Formalists (Hilbert) and Intuitionists (Brouwer). The formalist mathematicians would say that any point in our product space is either dangerous or not dangerous. But the intuitionist mathematicians would argue that *the Law of the Excluded Middle is not universally valid*. It cannot be decided whether a point is dangerous or not. From a realistic point of view, one might find that the intuitionists are merely right. Our conclusion would be, then, that the behaviour of the products of real world off-diagonal coefficients $(a.b)$ is, actually, *unpredictable*, if their values are to be chosen in the domain $1/4 < (a.b) < \infty$.
We have also seen that there exists another, "safe" domain, which fortunately is quite distinct from the dangerous area, namely $0 \leq (a.b) \leq 1/4$. The boundary

1/4 of this safe domain may even be defined with a sensible pinch of salt. And it is expected that no weird mathematical problems will show up for this area of interest.

The Hyperbolic Connection

A Trigonometric Connection has been found for the "dangerous" domain in the space of products of matrix coefficients. But, satisfactory as it is, we were rather interested in a similar theory for the "safe" domain $0 \leq a.b \leq 1/4$. It is clear that the function $1/a.b = 2 + 2.cos(\pi.h)$ cannot be used for this domain, not because of its properties for doubling or halving the angle, but because of the fact that its range is limited to $0 \leq 1/a.b \leq 4$. For the safe domain, a range $0 \leq a.b \leq 1/4$ or $4 \leq 1/a.b < \infty$ would be needed instead. Hence the question is: do there exist functions which have the same properties for doubling and halving their arguments as the trigonometric functions, but also have a quite different range? The answer is *yes*. These functions do indeed exist. They are called *Hyperbolic Functions*. As the name already suggests, hyperbolic functions are associated with an (orthogonal) hyperbola, in contrast with the trigonometric functions, which are associated with a circle.
Consider (the part on the right of) an orthogonal hyperbola, represented by:

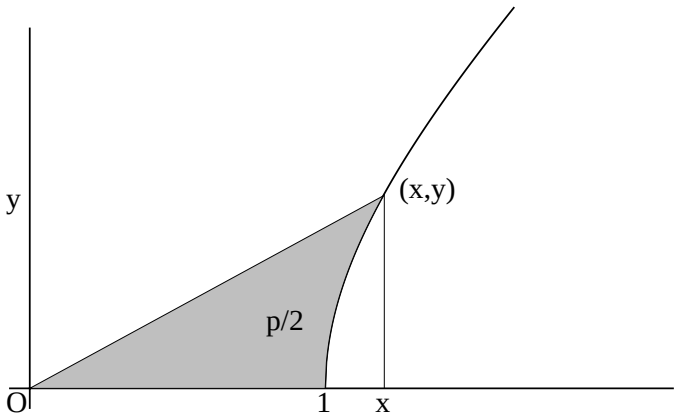
$$x^2 - y^2 = 1 \qquad \text{or} \qquad y(x) = \sqrt{x^2 - 1} \quad \text{for} \quad x \geq 1$$

An arbitrary area underneath $y(x)$ can be calculated with help of the integral: $\int_1^x y(t)\,dt$. Let's invoke a little help from MAPLE:

```
int{sqrt(t^2-1),t=1..x);
```

Then we find:

$$\int_1^x y(t)\,dt = \frac{1}{2}x\sqrt{x^2-1} - \frac{1}{2}ln\left(x + \sqrt{x^2-1}\right)$$



The first term in this expression is the area of a triangle with base x and height

y. The expression as a whole represents the area underneath y from 1 to x. Thus the second term in the expression represents the area which is spanned by the x-axis, a vector from the Origin to (x, y) and the curved line between $(1, 0)$ and (x, y). The latter area may be taken as an analogy with the area $\phi/2$ within a (unit-)circle sector. Hence we could make the following choice, when defining a "hyperbolic angle" $2.p/2$ instead of $2.\phi/2 = \phi$:

$$p = ln\left(x + \sqrt{x^2 - 1}\right) = ln(x + y)$$

Herewith we define a hyperbolic sine and a hyperbolic cosine:

$$sinh(p) = y \qquad cosh(p) = x$$

And a hyperbolic tangent, eventually:

$$tanh(p) = \frac{sinh(p)}{cosh(p)}$$

It is possible to find explicit expressions for the hyperbolic functions:

$$\begin{array}{rcrcl} & & p = ln(x + y) & \implies & x + y = e^{+p} \\ x^2 - y^2 = 1 & \implies & (x + y)(x - y) = 1 & \implies & x - y = e^{-p} \end{array}$$

Addition and subtraction, or solving the equations for x and y gives:

$$sinh(p) = \frac{e^{+p} - e^{-p}}{2} \qquad cosh(p) = \frac{e^{+p} + e^{-p}}{2}$$

And, eventually:

$$tanh(p) = \frac{e^{+p} - e^{-p}}{e^{+p} + e^{-p}}$$

The following formula is the hyperbolic analogy of $cos^2(x) + sin^2(x) = 1$:

$$cosh^2(p) - sinh^2(p) = 1 \quad \implies \quad cosh^2(p) - 1 = sinh^2(p)$$

Especially the function $cosh(p)$ will be of interest to us. It is immediately clear that the range of $cosh(p)$ is precisely the kind of completion which is needed for the "safe" domain of interest:

$$1 \leq cosh(p) < \infty \qquad \text{while} \qquad -1 \leq cos(\pi.h) \leq +1$$

Where the minimum value is attained as $cosh(0) = 1$. It is questioned now if formulas for doubling or halving the hyperbolic angles do indeed exist:

$$cosh(2.p) = \frac{e^{+2.p} + e^{-2.p}}{2} = \frac{\left(e^{+p}\right)^2 + \left(e^{-p}\right)^2 + 2.e^{+p}.e^{-p} - 2}{2} =$$

$$2\left(\frac{e^{+p}+e^{-p}}{2}\right)^2-1=2.cosh^2(p)-1 \quad \Longrightarrow$$

$$cosh(2.p)=2.cosh^2(p)-1 \quad \text{and} \quad cosh(\frac{1}{2}p)=\sqrt{\frac{1+cosh(p)}{2}}$$

Doubling the grid-spacing corresponds with doubling the hyperbolic angle, while halving the grid-spacing corresponds with halving the hyperbolic angle, meaning that the hyperbolic angle must be proportional to the grid-spacing:

$$p=Q.dx \quad \text{with} \quad Q\geq 0$$

The latter condition imposes no limitation on generality, because the hyperbolic cosine is symmetric: $cosh(p)=cosh(-p)$. Therefore the absolute value of its argument is all that matters. Thus Q is a positive constant which eventually has to be determined later on. The hyperbolic cosine is ready to take over now where the trigonometric cosine has failed. As with the trigonometric cosine, we shall propose:

$$a.b(dx)=\frac{1}{2+2.cosh(Q.dx)}$$

All persistent properties will remain the same, because they are only dependent upon the formulas for doubling and halving. And with respect to these formulas it makes no difference whether we use the trigonometric or hyperbolic connection. It may be even remarked that the trigonometric and the hyperbolic connection are transformed in each other at $(1/4, 1/4)$, by switching from an imaginary to a real argument, or vice versa, because:

$$cosh(j.\phi)=\frac{e^{+j.\phi}+e^{-j.\phi}}{2}=cos(\phi) \qquad \text{where} \quad j=\text{ imaginary unit}$$

The above formula for $a.b(dx)$ can be written in a more transparent form:

$$a.b(dx)=\frac{1}{2+2.cosh(Q.dx)}=\frac{1}{2+2.\left[2.cosh^2(Q/2.dx)-1\right]}$$

Resulting in:

$$a.b=\frac{1/4}{cosh^2(Q/2.dx)} \quad \Longrightarrow \quad \sqrt{a.b}=\frac{1/2}{cosh(Q/2.dx)}$$

From the preceding paragraph, we also have:

$$\frac{b}{a}=e^{P.dx} \quad \Longrightarrow \quad \sqrt{\frac{b}{a}}=e^{P/2.dx} \quad \text{and} \quad \sqrt{\frac{a}{b}}=e^{-P/2.dx}$$

It is a simple matter now to find the explicit formulas, relating each of the off-diagonal coefficients to the distances dx in the grid:

$$b = \sqrt{\frac{b}{a}}\sqrt{a.b} = \frac{\frac{1}{2}e^{+P/2.dx}}{cosh(Q/2.dx)} = \frac{e^{+P/2.dx}}{e^{+Q/2.dx} + e^{-Q/2.dx}}$$
$$a = \sqrt{\frac{a}{b}}\sqrt{a.b} = \frac{\frac{1}{2}e^{-P/2.dx}}{cosh(Q/2.dx)} = \frac{e^{-P/2.dx}}{e^{+Q/2.dx} + e^{-Q/2.dx}}$$

Herewith a new light is shed upon all kind of persistent properties. For example:

$$a + b = \frac{e^{+P/2.dx}}{e^{+Q/2.dx} + e^{-Q/2.dx}} + \frac{e^{-P/2.dx}}{e^{+Q/2.dx} + e^{-Q/2.dx}} =$$

$$\frac{e^{+P/2.dx} + e^{-P/2.dx}}{e^{+Q/2.dx} + e^{-Q/2.dx}} = \frac{cosh(P/2.dx)}{cosh(Q/2.dx)}$$

It is seen therefrom that the following relationships are true:

$$\begin{aligned} a + b < 1 \quad &\Longleftrightarrow \quad |P| < |Q| \\ a + b = 1 \quad &\Longleftrightarrow \quad |P| = |Q| \\ a + b > 1 \quad &\Longleftrightarrow \quad |P| > |Q| \end{aligned}$$

The absolute values come from: $cosh(-P.dx) = cosh(P.dx) = cosh(|P|.dx)$. Another interesting quantity is the so-called *matrix discriminant*, for which the *sign* was found to be persistent. Meanwhile the sign has even become positive, because of $a.b \leq 1/4$ and:

$$1 - 4.a.b = 1 - 4.\frac{1/4}{cosh^2(Q/2.dx)} = \frac{cosh^2(Q/2.dx) - 1}{cosh^2(Q/2.dx)}$$

$$= \frac{sinh^2(Q/2.dx)}{cosh^2(Q/2.dx)} = tanh^2(Q/2.dx)$$

Let's try now for even more general solutions of the finite difference equation which is accompanying the tri-diagonal system of equations:

$$-b.T_{i-1} + T_i - a.T_{i+1} = 0$$

Solutions of the form $T_i = K^{i-1}$ will be attempted. Substitution leads to:

$$K^{i-2}\left[-b + K - a.K^2\right] = 0 \quad \text{or} \quad a.K^2 - K + b = 0$$

A real-valued solution exists iff the discriminant $1 - 4.a.b \geq 0$. Since the discriminant equals $tanh^2(Q/2.dx)$, there is no question about it. Substitute now the expressions that we have found for a and b, giving:

$$e^{-P/2.dx}.K^2 - \left(e^{+Q/2.dx} + e^{-Q/2.dx}\right)K + e^{+P/2.dx} = 0$$

$$K^2 - e^{P/2.dx}\left(e^{+Q/2.dx} + e^{-Q/2.dx}\right)K + e^{P.dx} = 0$$

$$\left(K - e^{(P+Q)/2.dx}\right)\left(K - e^{(P-Q)/2.dx}\right) = 0$$

Resulting in:

$$T_i = \lambda.K_1^{i-1} + \mu.K_2^{i-1} \qquad \text{with} \quad K_1 = e^{(P+Q)/2.dx} \quad \text{and} \quad K_2 = e^{(P-Q)/2.dx}$$

At last, $(i-1).dx$ can be simply replaced by (x), yielding the equivalent:

$$T(x) = \lambda.e^{(P+Q)/2.x} + \mu.e^{(P-Q)/2.x}$$

Which, at the same time, is recognized as a general *Analytical* Solution.

Governing equation

Imagine a grid which is extremely fine. Actually, the grid is so fine that it can by no physical means be distinguished from a "true" continuum. Then it may be assumed that also the numerical solution at such a fine grid can by no means be distinguished from the analytical one. Imagine now that we use our Newton-Rhapson MultiGrid Solver for coarsening, step by step, this immensely fine grid. Also suppose that, at some step, the matrix coefficients correspond with some persistent scheme. If this is the case, only once, then the matrix-coefficients will *always* be changed in such a way that the scheme persists through all of the coarser and finer grids of the solver. However, an "exact" solution will be obtained by using a persistent scheme at a immensely fine grid. Coarsening does not change the persistent scheme, nor does it change the "exact" solution, as it is specified at the points of the successive grids. These considerations lead to the following Corollary:
Even if we use a persistent scheme at a very coarse grid, the solution will always be nothing else but just a specification of the exact solution at the grid points. To put it in a simple and straightforward way: any one-dimensional persistent scheme will give rise to numerical solutions which are **exact** at the grid points. Hence such a scheme in fact will be identical to an Exact or Analytical Scheme. Consider again the finite difference equation which is associated with a tri-diagonal system of equations:

$$-b.T_{i-1} + T_i - a.T_{i+1} = 0$$

Meanwhile, we have found explicit formulas for the coefficients a and b:

$$b = \frac{e^{+P/2.dx}}{e^{+Q/2.dx} + e^{-Q/2.dx}} \qquad a = \frac{e^{-P/2.dx}}{e^{+Q/2.dx} + e^{-Q/2.dx}}$$

Let's proceed now with the F.D. equation, working towards a continuous version:

$$-e^{+P/2.dx}.T(x-dx) + \left(e^{+Q/2.dx} + e^{-Q/2.dx}\right)T(x) - e^{-P/2.dx}.T(x+dx) = 0$$

The exponential functions are expanded into a Taylor series, while carefully preserving *all* terms of second order:

$$-\left[1+P/2.dx+(P/2.dx)^2/2\right]T(x-dx)$$

$$+\left[2+(Q/2.dx)^2\right]T(x)$$

$$-\left[1-P/2.dx+(P/2.dx)^2/2\right]T(x+dx)=0$$

Collect terms with respect to different powers of (dx):

$$-\left[T(x-dx)-2.T(x)+T(x+dx)\right]\ +\ dx.P\left[T(x+dx)-T(x-dx)\right]/2$$

$$-dx^2\left[(P/2)^2/2.T(x-dx)-(Q/2)^2.T(x)+(P/2)^2/2.T(x+dx)\right]=0$$

The next step is to divide everything by $(-dx^2)$ and to work out every term. First (order) derivative:

$$\frac{T(x+dx)-T(x-dx)}{2.dx}=\frac{dT}{dx}\quad\text{for}\quad dx\to 0$$

Second (order) derivative:

$$\frac{\left[T(x+dx)-T(x)\right]/dx-\left[T(x)-T(x-dx)\right]/dx}{dx}=$$

$$\frac{dT/dx|_{x+\frac{1}{2}dx}-dT/dx|_{x-\frac{1}{2}dx}}{dx}=\frac{d^2T}{dx^2}\quad\text{for}\quad dx\to 0$$

The second order derivative is also recognized in the term with dx^2 which, after division by $(-dx^2)$, has become:

$$\left[(P/2)^2/2.T(x-dx)-(Q/2)^2.T(x)+(P/2)^2/2.T(x+dx)\right]$$

$$=(P/2)^2\frac{1}{2}\left[T(x-dx)-2.T(x)+T(x+dx)\right]+\left[(P/2)^2-(Q/2)^2\right]T(x)=$$

$$=(P/2)^2\frac{dx^2}{2}\left[\frac{T(x-dx)-2.T(x)+T(x+dx)}{dx^2}\right]+\left[(P/2)^2-(Q/2)^2\right]T(x)$$

Yielding a more or less continuous representation:

$$\left[1+\frac{1}{2}(P/2.dx)^2\right]\frac{d^2T}{dx^2}-P\frac{dT}{dx}+\left[(P/2)^2-(Q/2)^2\right]T(x)=0$$

We have deliberately left in the term $(P/2.dx)^2$, though it will go to zero with $dx\to 0$. This term is commonly known as "false diffusion" and it should be emphasized here that such a false diffusion term is can *not* be neglected if one wants to preserve all terms of a second order approximation for the governing differential equation. (The term especially becomes important for large Péchlet

numbers $|P|$.)
It is noted that "false diffusion", as mentioned here, should be well distinguished from the false diffusion in the book by Patankar (1980). The term is used there in connection with directional dependence of convection in a 2-D grid, which is quite a different matter.
If we finally do $dx \rightarrow 0$, then we find for the governing differential equation:

$$\frac{d^2T}{dx^2} - P\frac{dT}{dx} + \frac{1}{4}(P^2 - Q^2)T(x) = 0$$

This linear ODE (Ordinary Differential Equation) can be solved by conventional or by less conventional mathematical means. Finding the characteristic equation belongs to the former category:

$$\lambda^2 - P.\lambda + \frac{1}{4}(P^2 - Q^2) = 0$$

Giving:

$$\left[\lambda - \frac{1}{2}(P+Q)\right]\left[\lambda - \frac{1}{2}(P-Q)\right] = 0 \qquad \Longrightarrow$$

$$\lambda = \frac{1}{2}(P+Q) \quad \text{or} \quad \lambda = \frac{1}{2}(P-Q)$$

And the accompanying solution:

$$T(x) = \lambda.e^{(P+Q)/2.x} + \mu.e^{(P-Q)/2.x}$$

Which has been found at an earlier stage.
The fact that oscillating (complex) solutions cannot be found in this manner is quite remarkable. This is in agreement with the finding that the discriminant of the characteristic equation is always positive. Write the differential equation in the form:

$$A\frac{d^2T}{dx^2} + B\frac{dT}{dx} + C.T = 0$$

Then we find for the discriminant of the characteristic equation:

$$(B/2A)^2 - (C/A) = (P/2)^2 - \frac{1}{4}(P^2 - Q^2) = \frac{1}{4}Q^2$$

$$\Longrightarrow \quad Q/2 = \sqrt{(B/2A)^2 - (C/A)}$$

The dependence on Q is also apparent in an expression that we have found for the discriminant of the accompanying tri-diagonal matrix:

$$1 - 4.a.b = tanh^2(Q/2.dx) \quad \Longrightarrow \quad \sqrt{1 - 4.a.b} = tanh\left[\sqrt{(B/2A)^2 - (C/A)}.dx\right]$$

Last but not least, the factor P is found to be equal to:

$$P/2 = -B/2.A$$

Upper and Lower case

There are two cases in which a tri-diagonal system may be reduced to a system with only two diagonals, resulting in either an Upper two-diagonal matrix U or a Lower two-diagonal matrix L:

$$\begin{bmatrix} . & . & . & & & & \\ & -b & 1 & 0 & & & \\ & & -b & 1 & 0 & & \\ & & & -b & 1 & 0 & \\ & & & & . & . & . \end{bmatrix} = L \qquad \begin{bmatrix} . & . & . & & & & \\ & 0 & 1 & -a & & & \\ & & 0 & 1 & -a & & \\ & & & 0 & 1 & -a & \\ & & & & . & . & . \end{bmatrix} = U$$

The properties of being an Upper or a Lower matrix are Persistent Properties. Use $(a', b') = (a^2, b^2)/(1 - 2.a.b)$ to prove:

$$a' = a^2 \quad \text{and} \quad b' = b = 0 \qquad \text{xor} \qquad b' = b^2 \quad \text{and} \quad a' = a = 0$$

$$a = \sqrt{a'} \quad \text{and} \quad b = b' = 0 \qquad \text{xor} \qquad b = \sqrt{b'} \quad \text{and} \quad a' = a = 0$$

Here "xor" stands for: eXclusive OR. We can also write:

$$\begin{aligned} a(2.dx) = a^2(dx) \qquad & \text{and} \qquad a(dx/2) = a^{1/2}(dx) \\ b(2.dx) = b^2(dx) \qquad & \text{and} \qquad b(dx/2) = b^{1/2}(dx) \end{aligned}$$

And subsequently develop another piece of theory, which will be analogous then to the theory of the "Quotient Function".
It's easy to find the Finite Difference equations which are associated with the Lower and Upper matrix, respectively:

$$\begin{aligned} -b.T_{i-1} + T_i = 0 \qquad & \text{xor} \qquad T_i - a.T_{i+1} = 0 \\ T_i = b.T_{i-1} \qquad & \text{xor} \qquad T_i = a.T_{i+1} \end{aligned}$$

But the equations can also be written as follows, assigning to i an increment or a decrement, as it is appropriate:

$$\begin{aligned} -b.T_i + T_{i+1} = 0 \qquad & \text{xor} \qquad T_{i-1} - a.T_i = 0 \\ T_i = 1/b.T_{i+1} \qquad & \text{xor} \qquad T_i = 1/a.T_{i-1} \end{aligned}$$

Herewith we see that an Upper matrix can be made equivalent to a Lower matrix, and vice versa, as follows:

$$\begin{bmatrix} . & . & . & & & & \\ & -b & 1 & 0 & & & \\ & & -b & 1 & 0 & & \\ & & & -b & 1 & 0 & \\ & & & & . & . & . \end{bmatrix} \equiv \begin{bmatrix} . & . & . & & & & \\ & 0 & 1 & -1/b & & & \\ & & 0 & 1 & -1/b & & \\ & & & 0 & 1 & -1/b & \\ & & & & . & . & . \end{bmatrix}$$

$$\begin{bmatrix} . & . & . & & & \\ & 0 & 1 & -a & & \\ & & 0 & 1 & -a & \\ & & & 0 & 1 & -a \\ & & & & . & . & . \end{bmatrix} \equiv \begin{bmatrix} . & . & . & & & \\ & -1/a & 1 & 0 & & \\ & & -1/a & 1 & 0 & \\ & & & -1/a & 1 & 0 \\ & & & & . & . & . \end{bmatrix}$$

This is quite analogous to $(b/a)^{-1} = (a/b)$ and traversing the grid in the reverse direction.
A natural boundary condition for the Lower matrix is at $(i = 1)$, because, while going from the top to the bottom, the Lower matrix corresponds with the following F.D. equations:

$$T_1 = T_1 \ ; \ -b.T_1 + T_2 = 0 \ ; \ ... \ ; \ -b.T_{i-1} + T_i = 0 \ ; \ ... \ ; \ -b.T_{N-1} + T_N = 0$$

The system can be solved directly, by *forward* substitution:

$$T_2 = b.T_1 \ ; \ T_3 = b.T_2 = b^2.T_1 \ ; \ ... \ ; \ T_i = b^{i-1}.T_1 \ ; \ ... \ ; \ T_N = b^{N-1}.T_1$$

A natural boundary condition for the Upper matrix is at $(i = N)$, because, while going from the bottom to the top, the Upper matrix corresponds with the following F.D. equations:

$$T_N = T_N \ ; \ T_{N-1} - a.T_N = 0 \ ; \ ... \ ; \ T_i - a.T_{i+1} = 0 \ ; \ ... \ ; \ T_1 - a.T_2 = 0$$

The system can be solved directly, by *backward* substitution:

$$T_{N-1} = a.T_N \ , \ T_{N-2} = a^2.T_N \ , \ ... \ , \ T_{N-i} = a^i.T_N \ , \ ... \ , \ T_1 = a^{N-1}.T_N$$

Herewith the general F.D solutions for the L/U matrices become:

$$T_i = T_1.b^{i-1} \qquad \text{xor} \qquad T_{N-i} = T_N.a^i$$

Rewrite the Upper solution:

$$T_{N-i} = T_N.a^i \quad \Longrightarrow \quad T_1 = T_{N-(N-1)} = T_N.a^{N-1}$$

$$\Longrightarrow \quad T_i = T_{N-(N-i)} = T_N.a^{N-i} = T_N.a^{N-1}.a^{1-i} = T_1.(1/a)^{i-1}$$

Rewrite the Lower solution:

$$T_i = T_1.b^{i-1} \quad \Longrightarrow \quad T_N = T_1.b^{N-1}$$

$$\Longrightarrow \quad T_{N-i} = T_1.b^{N-i-1} = T_1.b^{N-1}.b^{-i} = T_N.(1/b)^i$$

We conclude that there exists quite some resemblance between the Lower and the Upper solutions:

$$T_N.a^{N-i} = T_1.(1/a)^{i-1} \qquad \text{xor} \qquad T_1.b^{i-1} = T_N.(1/b)^i$$

Let L denote the total length of the mesh. (Herewith we can take care of the fact that real exponents preferably should be dimensionless.) Substituting $x = (i-1).dx/L$ then leads to accompanying Analytical solutions:

$$T(x) = T(0).b^{x/L} \quad \Longleftrightarrow \quad T(L-x) = T(L).b^{-x/L}$$
$$T(x) = T(0).a^{-x/L} \quad \Longleftrightarrow \quad T(L-x) = T(L).a^{x/L}$$

The boundary conditions should, of course, be consistent herewith:

$$T(L) = T(0).b \qquad \text{xor} \qquad T(0) = T(L).a$$

It can be demonstrated that the Rule of Positive Coefficients must remain valid for Upper and Lower matrices, irrespective of any laws for grid coarsening and grid refinement. Take a closer look at the Upper solution $T_i = b^{i-1}$. For negative b, the solution T_i would change sign with every increment of i. If $i-1$ was even, hence b^{i-1} positive, then i would be odd, hence b^i negative. But then $i+1$ would be even again and b^{i+1} positive. The solution would exhibit a strong oscillatory behaviour. Even worse, the period of the oscillations would be proportional to the grid-spacing. If a numerical solution is assumed to converge to a smooth analytical solution, then such an "unstable" behaviour clearly does not lead to the desired result. Thus unstable numerical solutions are commonly considered as being unacceptable. We conclude that acceptable solutions are only obtained for $b > 0$. Very much the same argument can be employed, in order to prove, once more, that $a > 0$.

L.U. Decomposition

The product of the off-diagonal coefficients in an Upper or a Lower two-diagonal matrix is always zero. Therefore it is always in the safe domain $0 \leq a.b \leq 1/4$. This means, that, apart from the Rule of Positive Coefficients, *there isn't any further restriction* on the magnitude of a and b. Now a Lower and an Upper matrix can always be multiplied with each other. The result will be a tri-diagonal matrix:

$$\begin{bmatrix} . & . & . & & & \\ & -b & 1 & 0 & & \\ & & -b & 1 & 0 & \\ & & & -b & 1 & 0 \\ & & & & . & . & . \end{bmatrix} \begin{bmatrix} . & . & . & & & \\ & 0 & 1 & -a & & \\ & & 0 & 1 & -a & \\ & & & 0 & 1 & -a \\ & & & & . & . & . \end{bmatrix} =$$

$$\begin{bmatrix} . & . & . & & & \\ & -b & 1+a.b & -a & & \\ & & -b & 1+a.b & -a & \\ & & & -b & 1+a.b & -a \\ & & & & . & . & . \end{bmatrix}$$

The reverse procedure of this is called an *L.U. decomposition*. It can be carried out in a way which is uniform for all rows of the tri-diagonal matrix, provided that the main diagonal is of the form $(1+a.b)$. Which means that the equations should *not* be normalized. One step of the pivoting process is shown below. Assume that row (i) has been pivoted successfully. It has resulted in a row $(0 \quad 1 \quad -a)$, which is stored in the Upper matrix, and a pivot $(-b)$, which is stored in the Lower matrix. Then row $(i+1)$ will be pivoted as follows:

$$\begin{bmatrix} . & . & . & & & & \\ & 0 & 1 & -a & & & \\ & & -b-(-b.1) & 1+a.b-(-b.-a) & -a & & \\ & & & -b & 1+a.b & -a & \\ & & & & . & . & . \end{bmatrix}$$

We see that the pivot is again $(-b)$ and that row $(i+1)$ again becomes equal to $(0 \quad 1 \quad -a)$. Repeating the same process for all rows will result in an Upper and a Lower matrix, called the L.U. decomposition. The product of the Lower and the Upper matrix is equivalent to the original.
We have seen that there exist Upper and Lower solutions:

$$T_N.a^{N-i} = T_1.(1/a)^{i-1} \qquad \text{xor} \qquad T_1.b^{i-1} = T_N.(1/b)^i$$

We will demonstrate now that any linear superposition of these solutions is also a solution of the accompanying tri-diagonal system.
Substitute:

$$T_i = \lambda.(1/a)^{i-1} + \mu.b^{i-1}$$

Into the F.D. equation:

$$-b.T_{i-1} + (1+a.b).T_i - a.T_{i+1}$$

Giving, indeed:

$$(1+a.b)\left[\lambda.(1/a)^{i-1} + \mu.b^{i-1}\right]$$

$$-b\left[\lambda.a.(1/a)^{i-1} + \mu.1/b.b^{i-1}\right] - a\left[\lambda.1/a.(1/a)^{i-1} + \mu.b.b^{i-1}\right] =$$

$$\lambda.(1/a)^{i-1}\left[-b.a + (1+a.b) - a/a\right] + \mu.b^{i-1}\left[-b/b + (1+a.b) - a.b\right] \equiv 0$$

We seek a relationship between the coefficients (a,b) in our tri-diagonal systems and the coefficients (a,b) in the Upper and Lower matrices, which are formed by an L.U. decomposition of the former. The situation may seem somewhat confusing, because the Lower and Upper matrices as such are normed, while the accompanying tri-diagonal matrix in this chapter is not. Before using results from "The Hyperbolic Connection", we should take appropriate measures. Divide (a,b) by the diagonal coefficient $(1+a.b)$ and then equate:

$$\frac{b}{1+a.b} = \frac{e^{+P/2.dx}}{e^{+Q/2.dx} + e^{-Q/2.dx}}$$

$$\frac{a}{1+a.b} = \frac{e^{-P/2.dx}}{e^{+Q/2.dx} + e^{-Q/2.dx}}$$

The product of the off-diagonal coefficients becomes:

$$\frac{b}{1+a.b}\frac{b}{1+a.b} = \frac{a.b}{(1+a.b)^2} = \frac{1}{\left(e^{+Q/2.dx}+e^{-Q/2.dx}\right)^2}$$

Check out safe and dangerous domains first:

$$\frac{a.b}{(1+a.b)^2} \leq \frac{1}{4} \quad \Longleftrightarrow \quad 4.a.b \leq 1+2.a.b+(a.b)^2$$

$$\Longleftrightarrow \quad (a.b)^2 - 2.a.b + 1 = (a.b-1)^2 \geq 0$$

Proving once more that the dangerous domain doesn't exist anymore for the newly defined coefficients (a, b). Now manipulate the right hand side, in such a way that it assumes the same form as the left hand side:

$$\frac{a.b}{(1+a.b)^2} = \frac{1}{\left(e^{+Q/2.dx}+e^{-Q/2.dx}\right)^2} = \frac{1}{e^{-Q.dx}\left(e^{+Q.dx}+1\right)^2} = \frac{e^{Q.dx}}{\left(1+e^{Q.dx}\right)^2}$$

We conclude therefrom that, for the newly defined coefficients:

$$a.b = e^{Q.dx} \qquad \text{and} \qquad \frac{b}{a} = e^{P.dx}$$

The latter result remains unchanged, namely.
Herewith we find new expressions for the off-diagonal coefficients in the L.U. decomposition:

$$b = \sqrt{\frac{b}{a}}.\sqrt{a.b} = e^{P/2.dx}e^{Q/2.dx} = e^{\frac{1}{2}(P+Q).dx}$$

$$a = \sqrt{\frac{a}{b}}.\sqrt{a.b} = e^{-P/2.dx}e^{Q/2.dx} = e^{-\frac{1}{2}(P-Q).dx}$$

For the Analytical Solution we find:

$$T(x) = \lambda.a^{-x} + \mu.b^{x} = \lambda.e^{\frac{1}{2}(P-Q).x} + \mu.e^{\frac{1}{2}(P+Q).x}$$

It is evident herewith that the superposition of solutions of the Upper and Lower equations, forming the L.U. decomposition of the tri-diagonal system, is indeed identical with the general solution of the tri-diagonal system itself, as it has been found in "The Hyperbolic Connection".
Last but not least, it should be questioned what the governing equations are, corresponding with the Upper and the Lower matrices. To that end, we could set up Taylor expansions for the coefficients a and b. But this would lead to finite difference schemes which are somewhat inconsistent with standard results from numerical analysis. This is the reason why the F.D. equations are modified as follows, before they serve as our starting point:

$$T_{i-1} - 1/b.T_i = 0 \qquad \text{xor} \qquad T_{i+1} - 1/a.T_i = 0$$

The coefficients $1/a$ and $1/b$ are developed into a Taylor series expansion. Only terms up and including the first order are retained:

$$1/b = e^{-\frac{1}{2}(P+Q).dx} \approx 1 - \frac{1}{2}(P+Q).dx$$
$$1/a = e^{+\frac{1}{2}(P-Q).dx} \approx 1 + \frac{1}{2}(P-Q).dx$$

Substitute in the Upper and Lower F.D. schemes. Then:

$$T_{i-1} - 1/b.T_i = T_{i-1} - \left[1 - \frac{1}{2}(P+Q).dx\right] T_i = 0$$

$$\iff \quad -\frac{T_i - T_{i-1}}{dx} + \frac{1}{2}(P+Q).T_i = 0 \qquad : \text{forward scheme } \rightarrow$$

$$T_{i+1} - 1/a.T_i = T_{i+1} - \left[1 + \frac{1}{2}(P-Q).dx\right] T_i = 0$$

$$\iff \quad \frac{T_{i+1} - T_i}{dx} - \frac{1}{2}(P-Q).T_i = 0 \qquad : \text{backward scheme } \leftarrow$$

Taking the limit for $dx \rightarrow 0$ results in:

$$\frac{dT}{dx} - \frac{1}{2}(P+Q).T = 0 \qquad \text{xor} \qquad \frac{dT}{dx} - \frac{1}{2}(P-Q).T = 0$$

Let $\alpha = (P+Q)/2$ and $\beta = (P-Q)/2$, then: $P = \alpha + \beta$ and $Q = \alpha - \beta$, thus assuring that α and β can be arbitrary real numbers. Hence, in the limit for $(dx \rightarrow 0)$, the backward and the forward schemes boil down to differential equations which are both of the same type, namely:

$$\frac{dT}{dx} - \gamma.T = 0 \qquad \text{with arbitrary } \gamma = \alpha, \beta$$

Note. It can be conceived that the L.U. decomposition of the tri-diagonal matrix into a Lower and an Upper matrix corresponds with a decomposition of the accompanying differential operator into a "lower" and an "upper" part:

$$\left[\frac{d}{dx} - \frac{1}{2}(P+Q)\right]\left[\frac{d}{dx} - \frac{1}{2}(P-Q)\right] = \frac{d^2}{dx^2} - P\frac{d}{dx} + \frac{1}{4}(P^2 - Q^2)$$

All Possible Cases

Having established a firm relationship between the tri-diagonal system and the governing ordinary differential equation, we should subsequently take care of all

possible (special) cases. There are eight of them, which can be enumerated as follows:

$$A \neq 0\ ,\ B \neq 0\ ,\ C \neq 0 \qquad (1)$$
$$A \neq 0\ ,\ B \neq 0\ ,\ C = 0 \qquad (2)$$
$$A \neq 0\ ,\ B = 0\ ,\ C \neq 0 \qquad (3)$$
$$A \neq 0\ ,\ B = 0\ ,\ C = 0 \qquad (4)$$
$$A = 0\ ,\ B \neq 0\ ,\ C \neq 0 \qquad (5)$$
$$A = 0\ ,\ B \neq 0\ ,\ C = 0 \qquad (6)$$
$$A = 0\ ,\ B = 0\ ,\ C \neq 0 \qquad (7)$$
$$A = 0\ ,\ B = 0\ ,\ C = 0 \qquad (8)$$

These cases are applicable to the governing equation:

$$A\frac{d^2T}{dx^2} + B\frac{dT}{dx} + C.T = 0$$

- The (1)'st case ($A \neq 0\ ,\ B \neq 0\ ,\ C \neq 0$) is the most general one and has been covered in detail in "Governing Equation". The only requirement here is that the discriminant $(B/2.A)^2 - (C/A)$ shall not be negative.
- The (2)'nd case ($A \neq 0\ ,\ B \neq 0\ ,\ C = 0$) corresponds with the governing equation:

$$A\frac{d^2T}{dx^2} + B\frac{dT}{dx} = 0 \quad \Longleftrightarrow \quad -\frac{d^2T}{dx^2} + P\frac{dT}{dx} = 0$$

The differential equation for Convection and Diffusion is recognized.
From ($P^2 - Q^2 = C = 0$) it follows that: $|P| = |Q|$. But then, from "The Hyperbolic Connection", we know:

$$a + b = 1 \quad \Longleftrightarrow \quad |P| = |Q|$$

The case ($a + b = 1$) has been covered at length in "Some Stable Solutions" and a solution of the ODE for Convection and Diffusion has been found there also.
- The (3)'rd case is ($A \neq 0\ ,\ B = 0\ ,\ C \neq 0$). The governing equation is:

$$A\frac{d^2T}{dx^2} + C.T = 0 \quad \Longleftrightarrow \quad \frac{d^2T}{dx^2} - \frac{1}{4}Q^2T = 0$$

Meaning that $P = 0$, hence $a = b$. And the discriminant is $Q^2 = -C/A$. A safe solution of the tri-diagonal system can only exists if the latter is a positive real number. Again, oscillatory and vibrating solutions *cannot* be (safely) obtained with a tri-diagonal system of linear equations.
- The (4)'th case is ($A \neq 0\ ,\ B = 0\ ,\ C = 0$). The governing equation is:

$$\frac{d^2T}{dx^2} = 0$$

Meaning that $P = 0$, hence $a = b$. And $P^2 = Q^2$, hence $Q = 0$. Therefore $a.b = 1/4$. This is only possible if: $a = b = 1/2$. The tri-diagonal system corresponds with the Finite Difference scheme for Pure 1-D Diffusion:

$$-\frac{1}{2}T_{i-1} + T_i - \frac{1}{2}T_{i+1} = 0$$

It is recognized that this is equivalent with one of the special cases, as noted in "Persistent Schemes":

$$a = b = \frac{1}{2} \quad \text{Symmetric matrix}$$

The solution of both the governing equation and the tri-diagonal system is a straight line between the boundaries, as has been established in the preamble of "Some Stable Solutions".
- The (5)'th case is ($A = 0$, $B \neq 0$, $C \neq 0$). The governing equation is:

$$B\frac{dT}{dx} + C.T = 0$$

It was proved in "L.U. Decomposition" that the governing equations of the Upper and Lower matrices are of the same type, namely:

$$\frac{dT}{dx} - \gamma.T = 0$$

Details are handled at length in the abovementioned chapter.
Here, of course: $\gamma = -C/B$.
- The (6)'th case is ($A = 0$, $B \neq 0$, $C = 0$). The governing equation is:

$$\frac{dT}{dx} = 0$$

This case is recognized as a further specialization of (5), though it was already mentioned as such in "Persistent Schemes":

$$a = 0, b = 1 \quad \text{Lower diagonal matrix}$$
$$a = 1, b = 0 \quad \text{Upper diagonal matrix}$$

- The (7)'th case is ($A = 0$, $B = 0$, $C \neq 0$). The governing equation is:

$$T = 0$$

Equivalent again with one of the special cases in "Persistent Schemes":

$$a = b = 0 \quad \text{Identity matrix}$$

- The (8)'h case is ($A = 0$, $B = 0$, $C = 0$). The ultimate degenerate case. There is no governing equation and no tri-diagonal system of equations at all.

The Main Result

Consider a one-dimensional uniform mesh with grid-spacing dx. Neighbouring grid-points in the mesh are coupled by coefficients a in forward direction and by coefficients b in backward direction. Setting up linear equations for such a grid gives rise to a tri-diagonal matrix, which can be (re-)normalized to obtain 1's on the main diagonal:

$$\begin{bmatrix} . & . & . & & & & \\ & -b & 1 & -a & & & \\ & & -b & 1 & -a & & \\ & & & -b & 1 & -a & \\ & & & & . & . & . \end{bmatrix}$$

The system of equations can be solved by employing a Newton-Rhapson Multi-Grid method. By employing the requirement that Properties of the tri-diagonal system should be *Persistent* on any coarsened or refined grid, we find that the Rule of Positive Coefficients is universally valid. And the *discriminant* of the equations system must be positive:

$$a \geq 0 \quad \text{and} \quad b \geq 0 \quad \text{and} \quad 1 - 4.a.b \geq 0$$

In the limiting case of a immensely fine grid, the system of equations becomes associated with a linear ODE of second order, called the Governing Equation:

$$A\frac{d^2T}{dx^2} + B\frac{dT}{dx} + C.T = 0$$

Where x = coordinate, A, B, C = constants and $T(x)$ = solution.
It is conjectured that the *discriminant* of the characteristic equation of the ODE must be positive (or zero). Having investigated all special cases, this turns out to be the *only* condition:

$$B^2 - 4.A.C \geq 0$$

Hence, oscillatory solutions can *never* be described by the governing ODE of a persistent tri-diagonal system of equations.
The coefficients a and b can be expressed in the coefficients A, B, C of the governing ODE and the grid-spacing dx. The solution of the tri-diagonal system is nothing else but a sampling on the grid of the Analytical Solution belonging to the governing ODE.

P.P. Summary

The coefficients of a tri-diagonal matrix, when associated with a 1-D uniform mesh, exhibit Persistent Properties. By definition, such P.P. are independent of the grid-spacing, while using a mesh refinement or coarsening procedure.
A (non exhaustive and sometimes redundant) list of Persistent Properties has

been produced below. If appropriate, the coefficients associated with the coarser grid are indicated with a prime accent $'$.

$$a = 0$$
$$b = 0$$

$$a > 0$$
$$b > 0$$

$$\begin{aligned} a = 0 \quad &\text{and} \quad b = 0 \\ a = 1/2 \quad &\text{and} \quad b = 1/2 \\ a = 0 \quad &\text{and} \quad b = 1 \\ a = 1 \quad &\text{and} \quad b = 0 \end{aligned}$$

$$\begin{aligned} a' + b' < a + b < 1 \qquad (*) \\ a' + b' = a + b = 1 \qquad (*) \\ a' + b' > a + b > 1 \qquad (*) \end{aligned}$$

$$\begin{aligned} a < b \quad &\text{or} \quad a'/b' < a/b < 1 \\ a = b \quad &\text{or} \quad a'/b' = a/b = 1 \\ a > b \quad &\text{or} \quad a'/b' > a/b > 1 \end{aligned}$$

$$a'.b' = a.b = 0$$
$$\begin{aligned} 1 - 4.a.b > 0 \quad &\text{or} \quad a'.b' < a.b < 1/4 \\ 1 - 4.a.b = 0 \quad &\text{or} \quad a'.b' = a.b = 1/4 \\ 1 - 4.a.b < 0 \quad &\text{or} \quad a'.b' > a.b > 1/4 \end{aligned}$$
$$a'.b' = a.b = 1$$

$$\begin{aligned} a'.b' = \tfrac{3}{2} + \tfrac{\sqrt{5}}{2} \quad &\Longleftrightarrow \quad a.b = \tfrac{3}{2} - \tfrac{\sqrt{5}}{2} \\ a'.b' = \tfrac{3}{2} - \tfrac{\sqrt{5}}{2} \quad &\Longleftrightarrow \quad a.b = \tfrac{3}{2} + \tfrac{\sqrt{5}}{2} \end{aligned}$$

Note (*). We didn't actually carry out a complete proof for these statements. Here comes:

$$a' + b' = \frac{a^2 + b^2}{1 - 2.a.b} < a + b \quad \Longleftrightarrow \quad a^2 + b^2 - (a + b)(1 - 2.a.b) < 0 \quad \Longleftrightarrow$$

$$a^2 + b^2 + 2.a^2.b + 2.a.b^2 - a - b < 0 \quad \Longleftrightarrow \quad (a + b - 1)(a + b + 2.a.b) < 0$$

$$\Longleftrightarrow \quad a+b<1 \qquad \text{because} \quad a+b+2.a.b>0$$

Then replace $<$ by $=$ and $>$ and repeat the sequence of arguments.

Reference

S.V. Patankar, "Numerical Heat Transfer and Fluid Flow",
Hemisphere PublishingCompany U.S.A. 1980.

APPENDIX I

Incremental Jacobi method

A well known iterative method for solving linear equations is easily derived by examining each of the n equations in the linear system $A.w = b$ in isolation. If in the i'th equation

$$\sum_{j=1}^{n} a_{i,j}.w_j = b_i$$

we solve for the value of w_i, while assuming the other entries of w remain fixed, we obtain:

$$w_i = (b_i - \sum_{j \neq i} a_{i,j}.w_j)/a_{i,i}$$

This suggests an iterative method defined by:

$$w_i^{(k)} = b_i - \sum_{j \neq i} a_{i,j}/a_{i,i}.w_j^{(k-1)}$$

Equivalently be written as:

$$w_i^{(k)} = w_i^{(k-1)} + b_i - \sum_{j=1}^{n} a_{i,j}/a_{i,i}.w_j^{(k-1)}$$

which is the Jacobi method. It will be assumed in the sequel that the equations system A is *always* normed, which means (in Pascal):

```
for i := 1 to N do
  d := 1/a[i,i];
for j := 1 to N do
  a[i,j] := a[i,j] * d;
```

A necessary condition being that the main diagonal of A is always non-zero. The main diagonal of the normed equations, hence, will be unity everywhere. The pseudo-code of the Jacobi method is then given by:

```
Initialize:
  w := 0
Iterations:
  w := w + (b - A.w)
```

Alternatively, an iterative method for solving the equations system $A.w = b$ may be devised by considering the fact that A can be written as $I - M$ and therefore we can probably use the sum of the Geometric Series:

$$A^{-1} = \frac{I}{I - M} = I + M + M^2 + M^3 + M^4 + M^5 + ...$$

Herewith it is assumed that the matrices M^n tend to become zero for large values of n. If such is the case, then the solution w can be computed by:

$$w = A^{-1}.b = \frac{I}{I-M}.b = (I + M + M^2 + M^3 + M^4 + M^5 + ...).b$$

The (pseudo)code of the above method, substituting $M = (I - A)$, is given by:

```
Initialize:
  r := b
  w := r
Iterations:
  r := r - A.r
  w := w + r
```

The difference with the standard Jacobi method is that iterations are performed now upon the *residual* r instead of the unknown solution w. Indeed r is the solution of $A.r = b$ for $b = 0$, hence the iterations are according to standard Jacobi on r. Next, the solution is incremented with the new residual found, and the whole is assumed to converge. Hence it's quite sensible to refer to such an iterative process as an *incremental Jacobi* method.
Let's formulate the requirement that M^n tends to become zero more precisely now. Any matrix M can be written as:

$$M = U^{-1}.\Lambda.U \quad \text{where} \quad \Lambda = \begin{bmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \lambda_3 & & \\ & & & . & \\ & & & & \lambda_n \end{bmatrix}$$

It follows that:

$$M^k = (U^{-1}.\Lambda.U)(U^{-1}.\Lambda.U)(U^{-1}.\Lambda.U)...(U^{-1}.\Lambda.U) = U^{-1}.\Lambda^k.U$$

Where:

$$\Lambda^k = \begin{bmatrix} \lambda_1^k & & & & \\ & \lambda_2^k & & & \\ & & \lambda_3^k & & \\ & & & . & \\ & & & & \lambda_n^k \end{bmatrix}$$

If the absolute values of all eigenvalues are smaller than 1, then:

$$lim_{k\to\infty}\lambda_i^k = 0$$

Then the transformation matrix U ant its inverse are multiplied by the zero matrix. Consequently also the result M^k must be zero. It thus seems that a

necessary and sufficient condition for the iteration process to converge is that the absolute values of all eigenvalues of the matrix M be smaller than 1.
Convergence of the method can be judged further by Gersgorin's Circle Theorem, which states the following. Define the radius R_i of the matrix-row (i) by:

$$R_i = \sum_{j \neq i} |m_{i,j}|$$

Then each (complex) Eigenvalue λ of the matrix M is in at least one of the following disks in the complex plane:

$$\{\lambda : |\lambda - m_{i,i}| \leq R_i\}$$

Now the diagonal elements of the iteration matrix M are all equal to zero, because we assumed the matrix A to be normalized and $M = I - A$ by definition. Furthermore, we find that all off-diagonal elements of M are equal to $a_{i,j}/a_{i,i}$. This means that a sufficient condition for all eigenvalues of the iteration matrix be less than 1 is:

$$|\lambda| \leq \sum_{j \neq i} |a_{i,j}|/|a_{i,i}| < 1 \qquad \Longleftrightarrow \qquad |a_{i,i}| > \sum_{j \neq i} |a_{i,j}|$$

Meaning that the original matrix A should better be *diagonally dominant*.
With the incremental Jacobi method in mind, several variations on the theme can be easily thought of, such as an incremental Gauss-Seidel and an incremental Successive OverRelaxation (SOR) Method. Or an incremental Jacobi method with Preconditioning. The latter two actually have been implemented by the author, for a 3-D problem which is related to the Solar Wind. See:

`http://huizen.dto.tudelft.nl/deBruijn/programs/zonwind.htm`

APPENDIX II

```
program MultiGrid;
{
  The End of MultiGrid
  ====================
  but only for 1-D ...

  Let the system of equations be given by  S.w = b .
  Iterate:  M := I - T.S ; b := (I + M).b ; S := I - M^2 .
  The background of this being (I - M)^(-1) = (I + M)/(I - M^2).
  The matrix  T  is a preconditioner, producing normed equations.
  For tri-diagonal 1-D systems, the matrix (I - M^2) recursively
  reduces to blocks around the main diagonal. It finally becomes
  a diagonal matrix which is normed, giving the Exact solution.

  CopyLefted by: Han de Bruijn (HdB)
}
const
  NN = 11 ; { Number of unknowns }

var
  e : array[1..2,1..2] of double;
  s,t : array[1..NN,0..2] of double;
  b,p : array[1..NN] of double;
  nr,no : array[1..NN] of byte;
  effe : text; { LOG file }
  L : byte ; getal : double ;

procedure Element(eps : double);
{
  Define Finite Element matrix
  ---------------------------- }
var
  a,b : double;

begin
{
  With Safety Condition:

  a.b = 0.5*(1 + eps)*0.5*(1 - eps)
      = (1 - sqr(eps))/4 < 1/4

  Because: eps = getal = Random < 1
}
  a := 0.5*(1 - eps);
  b := 0.5*(1 + eps);
```

```
  e[1,1] := +a ; e[1,2] := -a;
  e[2,1] := -b ; e[2,2] := +b;
end;

procedure Boekhouden(eerst : boolean);
{
  Administration
}
var
  i : byte;

begin
  if eerst then
  for i := 1 to NN do
    nr[i] := i;
{ Inverse: }
  for i := 1 to NN do
    no[nr[i]] := i;
end;

procedure Normeren;
{
  Make diagonal = unity
}
var
  i : byte;
  d : double;

begin
  for i := 1 to NN do
  begin
    d := 1/s[i,1];
    s[i,0] := s[i,0] * d;
    s[i,2] := s[i,2] * d;
    s[i,1] := 1;
    b[i] := b[i] * d;
  end;
end;

procedure Schoonmaken;
{
  Clear global matrix and vector
}
var
  i : byte;

begin
```

```
  for i := 1 to NN do
  begin
    s[i,0] := 0 ;
    s[i,1] := 0 ;
    s[i,2] := 0 ;
    b[i] := 0;
  end;
end;

procedure A_Symmetrisch;
{
  Fill global matrix & vector
}
var
  n,i,j,ii,jj : byte;

begin
  for n := 1 to NN-1 do
  begin
    for i := 1 to 2 do
    for j := 1 to 2 do
    begin
      ii := n+i-1 ; jj := j-i+1 ;
      s[ii,jj] := s[ii,jj] + e[i,j];
    end;
  end;

  b[1] := 1;
  s[1,1] := 1 ; s[1,2] := 0 ;

  b[NN] := 0;
  s[NN,0] := 0 ; s[NN,1] := 1 ;
end;

procedure Oplossen;
{
  Decomposition of tri-diagonal System
  ------------------------------------ }
var
  k : integer;
  diag, pivot : extended;

begin
  for k := 1 to NN do
  begin
    diag := s[k,1];
    if diag = 0 then begin
```

```
      Writeln('Oplossen: 0 on diagonal at: ',k);
      Halt;
    end ;
    pivot := s[k+1,0]/diag ;
    s[k+1,0] := pivot;
    if pivot = 0 then Continue ;
    s[k+1,1] := s[k+1,1] - pivot * s[k,2] ;
  end;
end;

procedure Oprollen;
{
  Solution of tri-diagonal System
  ------------------------------- }
var
  k : integer;
  diag, pivot : extended;

begin
  for k := 1 to NN do
  begin
    pivot := s[k+1,0] ;
    if pivot = 0 then Continue ;
    b[k+1] := b[k+1] - pivot * b[k] ;
  end;

  for k := NN downto 1 do
  begin
    pivot := b[k];
    diag := s[k,1];
    if diag = 0 then begin
      Writeln('Oplossen: 0 on diagonal at: ',k);
      Halt;
    end ;
    pivot := pivot - s[k,2] * b[k+1] ;
    b[k] := pivot/diag;
  end;
end;

procedure Vullen(term : double);
{
  Define System of Equations
  -------------------------- }
var
  k : integer;

begin
```

```
{ Bulk Assembly: }
  for k := 2 to NN-1 do
  begin
    s[k,1] := 1 ; b[k] := 0;
    s[k,0] := - 0.5*(1 + term);
    s[k,2] := - 0.5*(1 - term);
  end;
{ Left boundary: }
  s[1,1] := 1;
  s[1,2] := 0;
  b[1] := 1;
{ Right boundary: }
  s[NN,1] := 1;
  s[NN,0] := 0;
  b[NN] := 0;
end;

procedure Afdrukken;
{
  Print out Solution
}
var
  i : byte;

begin
  for i := 1 to NN do
    Write(b[no[i]]:7:4);
  Writeln;
end;

procedure Bekijken(onder,boven : byte);
{
  Take a snapshot in the LOG
}
var
  i : byte;

begin
  for i := onder+1 to boven do
    Writeln(effe,s[i,0]:9:5,
              ' ',s[i,1]:9:5,
              ' ',s[i,2]:9:5);
  for i := onder+1 to boven do
    Write(effe,b[i]:9:5,' ');
  Writeln(effe) ; Writeln(effe);
end;
```

```
procedure Newton(onder,boven : byte);
{
  Newton-Rhapson Multigrid
  ------------------------ }
var
  i,ii : byte;
  midden : byte;
  d : double;

begin
{
  Renormalization
}
  for i := onder+1 to boven do
  begin
    d := 1/s[i,1];
    s[i,0] := s[i,0] * d;
    s[i,2] := s[i,2] * d;
    s[i,1] := 1;
    b[i] := b[i] * d;
  end;
  Bekijken(onder,boven);

  if (onder+1 = boven) then Exit; { We are done ! }
  {
    M := I - S ; b := (I + M) * b
  }
  p := b;
  for i := onder+1 to boven do
  begin
    p[i] := b[i] - s[i,0]*b[i-1] - s[i,2]*b[i+1];
  end;
  b := p;
  {
    (I + M).(I - M) = I - M^2 -> main diagonal
  }
  for i := onder+1 to boven do
    s[i,1] := 1 - s[i,0]*s[i-1,2] - s[i,2]*s[i+1,0];
  {
    (I + M).(I - M) = I - M^2 -> off diagonal terms
  }
{ Two cases: # unknowns even or odd }
  if ((boven - onder) mod 2) = 0 then
    midden := onder + ((boven - onder) div 2);
  if ((boven - onder) mod 2) = 1 then
    midden := onder + ((boven - onder + 1) div 2);
```

```
{ Block out Odd indices: }
  for i := onder+1 to midden do
  begin
    ii := onder + 2*(i-onder) - 1;
    t[i,1] := s[ii,1];
    t[i,0] := - s[ii,0]*s[ii-1,0];
    t[i,2] := - s[ii,2]*s[ii+1,2];
    p[i] := b[ii];
    no[i] := nr[ii];
  end;
  t[onder+1,0] := 0;
  t[midden,2] := 0;

{ Block out Even indices: }
  for i := midden+1 to boven do
  begin
    ii := onder + 2*(i-midden);
    t[i,1] := s[ii,1];
    t[i,0] := - s[ii,0]*s[ii-1,0];
    t[i,2] := - s[ii,2]*s[ii+1,2];
    p[i] := b[ii];
    no[i] := nr[ii];
  end;
  t[midden+1,0] := 0;
  t[boven,2] := 0;

{ Recursively: }
  s := t ; b := p ;
  nr := no ;
  Newton(onder,midden);
  Newton(midden,boven);
end;

procedure FDM(eps : double);
{
  Finite Difference Method
  ------------------------ }
begin
{ Conventional }
  Boekhouden(true);
  Vullen(eps);
  Oplossen;
  Oprollen;
  Afdrukken;

{ Newton-Rhapson }
  Vullen(eps);
```

```
  Newton(0,NN);
  Boekhouden(false);
  Afdrukken;
end;

procedure FEM(eps : double);
{
  Finite Element Method
  ---------------------- }
begin
{ Conventional }
  Boekhouden(true);
  Element(eps);
  Schoonmaken;
  A_Symmetrisch;
  Oplossen;
  Oprollen;
  Afdrukken;

{ Newton-Rhapson }
  Schoonmaken;
  A_Symmetrisch;
  Newton(0,NN);
  Boekhouden(false);
  Afdrukken;
end;

begin
{
  The method should work for any a-symmetric,
  tri-diagonal and "safe" system of equations
}
  Assign(effe,'newton.log');
  Rewrite(effe);

  for L := 1 to 4 do
  begin
    Writeln;
    getal := Random;
    FDM(getal);
    FEM(getal);
  end;

  Close(effe);
end.
```