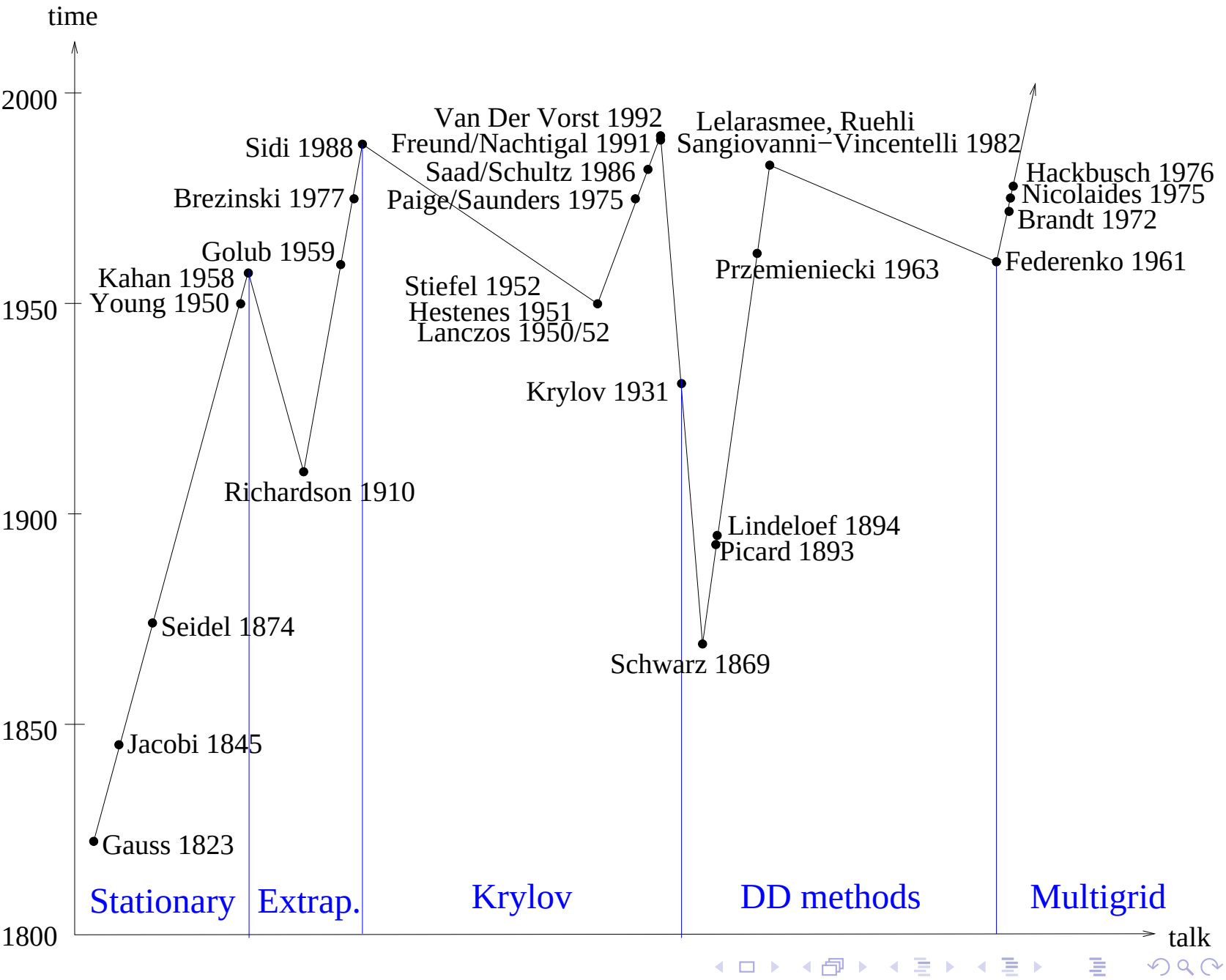
Extrapolation and Krylov Subspace Methods a Historical Approach

Martin J. Gander
martin.gander@unige.ch

University of Geneva

July 2012

Iterative Methods over Two Centuries



Gauss Invents an Iterative Method in a Letter

Gauss (1823), in a letter to Gerling: in order to compute a least squares solution based on angle measurements between the locations Berger Warte, Johannisberg, Taufstein and Milseburg:

Die Bedingungsgleichungen sind also:

$$
\begin{aligned}
0 &= +\quad 6 + 67a - 13b - 28c - 26d \\
0 &= -\ 7558 - 13a + 69b - 50c - 6d \\
0 &= -14604 - 28a - 50b + 156c - 78d \\
0 &= +22156 - 26a - 6b - 78c + 110d;
\end{aligned}
$$

$$\text{Summe} = 0.$$

Um nun indirect zu eliminiren, bemerke ich, dass, wenn 3 der Grössen a, b, c, d gleich 0 gesetzt werden, die vierte den grössten Werth bekommt, wenn d dafür gewählt wird. Natürlich muss jede Grösse aus ihrer eigenen Gleichung, also d aus der vierten, bestimmt werden. Ich setze also $d = -201$ und substituire diesen Werth. Die absoluten Theile werden dann: $+5232$, -6352, $+1074$, $+46$; das Übrige bleibt dasselbe.

Gauss' Method

Jetzt lasse ich b an die Reihe kommen, finde $b = +92$, substituire und finde die absoluten Theile: $+4036$, -4, -3526, -506. So fahre ich fort, bis nichts mehr zu corrigiren ist. Von dieser ganzen Rechnung schreibe ich aber in der Wirklichkeit bloss folgendes Schema:

	$d = -201$	$b = +92$	$a = -60$	$c = +12$	$a = +5$	$b = -2$	$a = -1$
$+6$	$+5232$	$+4036$	$+16$	-320	$+15$	$+41$	-26
-7558	-6352	-4	$+776$	$+176$	$+111$	-27	-14
-14604	$+1074$	-3526	-1846	$+26$	-114	-14	$+14$
$+22156$	$+46$	-506	$+1054$	$+118$	-12	0	$+26.$

Insofern ich die Rechnung nur auf das nächste 2000^{tel} [der] Secunde führe, sehe ich, dass jetzt nichts mehr zu corrigiren ist. Ich sammle daher

$a = -60$	$b = +92$	$c = +12$	$d = -201$
$+5$	-2		
-1			
-56	$+90$	$+12$	-201

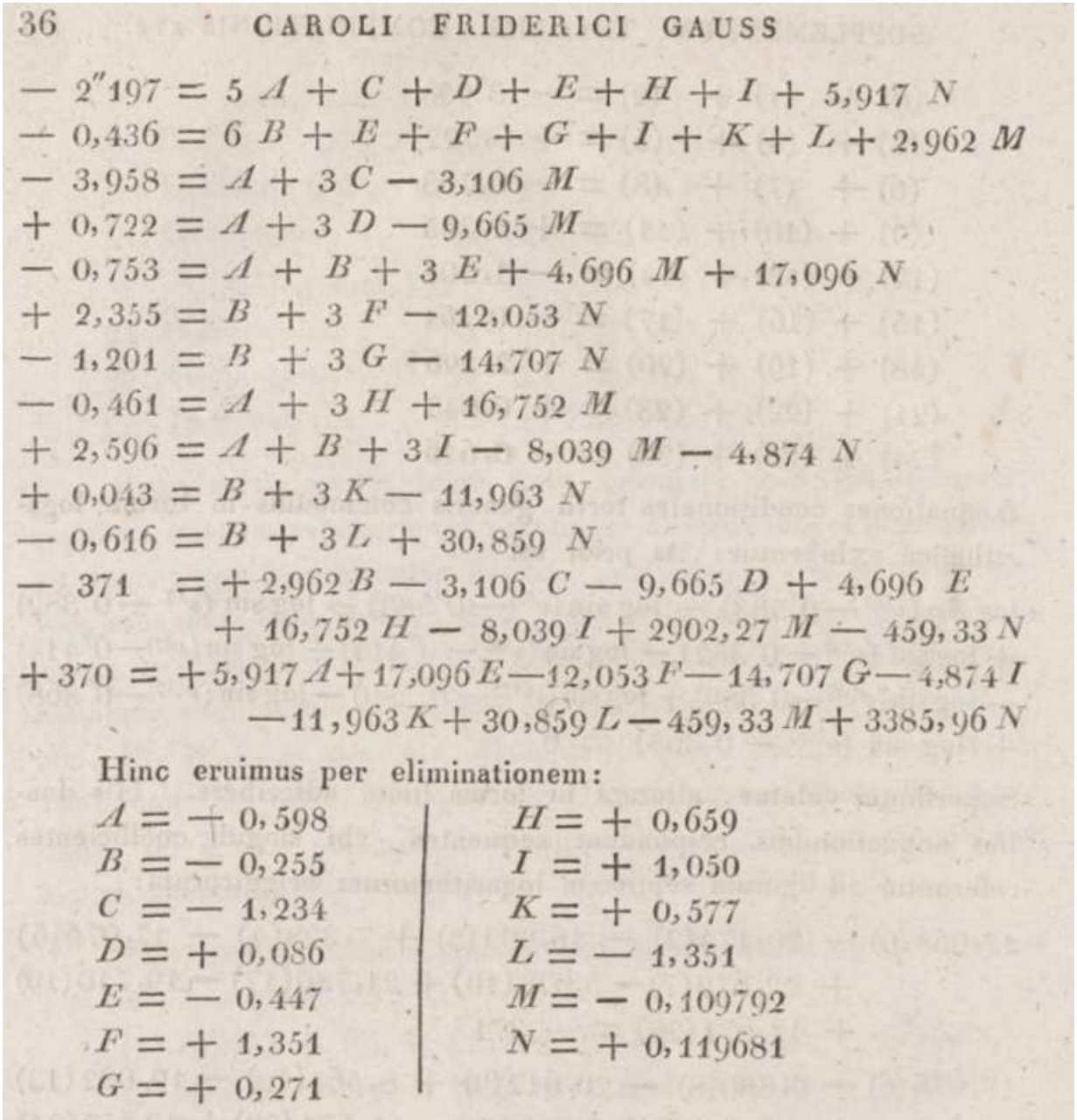
Calculations Make Happy !

Gauss concludes his letter with the statement:

Fast jeden Abend mache ich eine neue Auflage des Tableaus, wo immer leicht nachzuhelfen ist. Bei der Einförmigkeit des Messungsgeschäfts gibt dies immer eine angenehme Unterhaltung; man sieht dann auch immer gleich, ob etwas zweifelhaftes eingeschlichen ist, was noch wünschenswerth bleibt, etc. Ich empfehle Ihnen diesen Modus zur Nachahmung. Schwerlich werden Sie je wieder direct eliminiren, wenigstens nicht, wenn Sie mehr als 2 Unbekannte haben. Das indirecte Verfahren lässt sich halb im Schlafe ausführen, oder man kann während desselben an andere Dinge denken.

"... You will in future hardly eliminate directly, at least not when you have more than two unknowns. The indirect procedure can be done while one is half asleep, or is thinking about other things."

Supp. Theoria Combinationis Observationum 1828



36 CAROLI FRIDERICI GAUSS

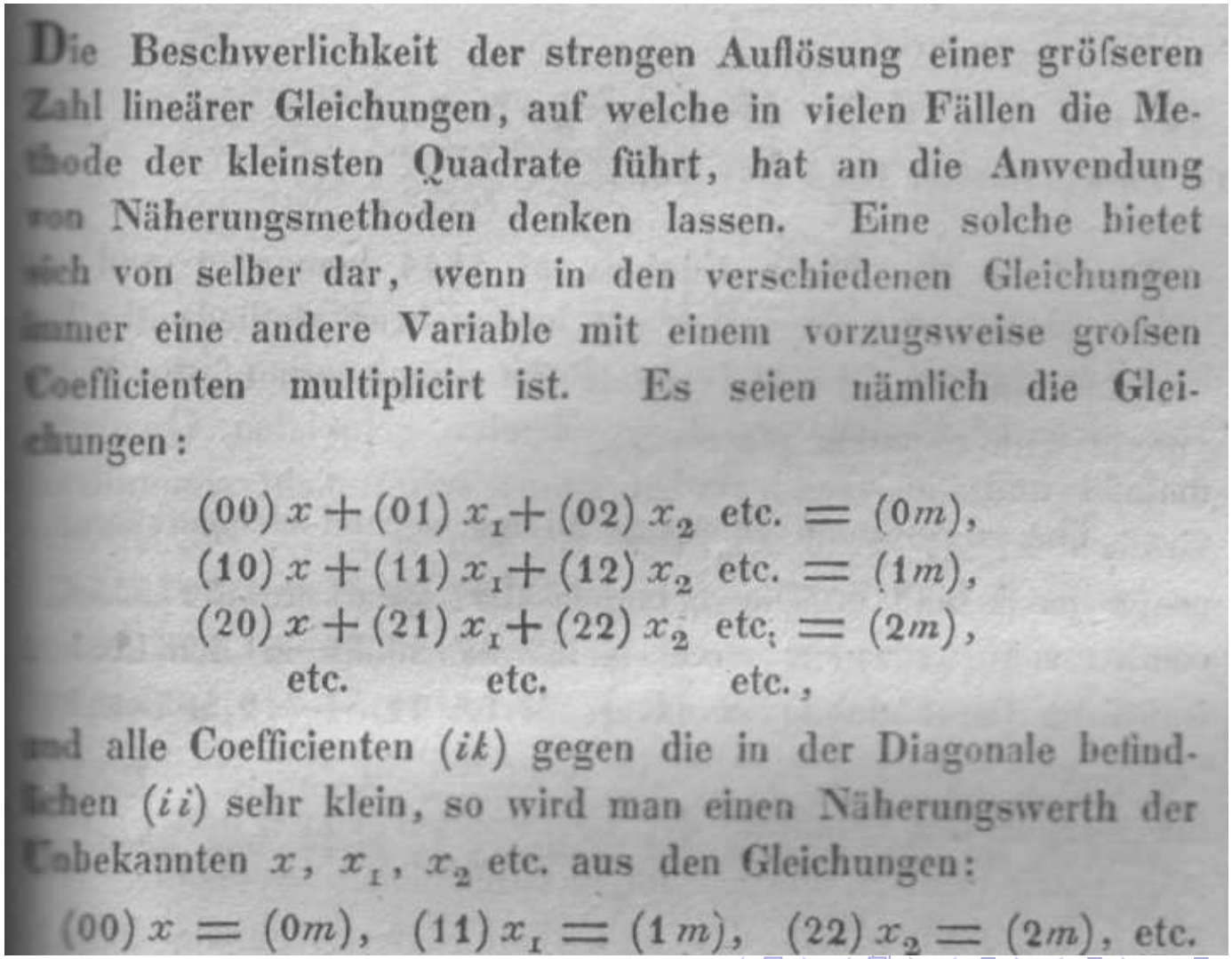
$$
\begin{aligned}
-2''197 &= 5A + C + D + E + H + I + 5{,}917\,N \\
-0{,}436 &= 6B + E + F + G + I + K + L + 2{,}962\,M \\
-3{,}958 &= A + 3C - 3{,}106\,M \\
+0{,}722 &= A + 3D - 9{,}665\,M \\
-0{,}753 &= A + B + 3E + 4{,}696\,M + 17{,}096\,N \\
+2{,}355 &= B + 3F - 12{,}053\,N \\
-1{,}201 &= B + 3G - 14{,}707\,N \\
-0{,}461 &= A + 3H + 16{,}752\,M \\
+2{,}596 &= A + B + 3I - 8{,}039\,M - 4{,}874\,N \\
+0{,}043 &= B + 3K - 11{,}963\,N \\
-0{,}616 &= B + 3L + 30{,}859\,N \\
-371 &= +2{,}962\,B - 3{,}106\,C - 9{,}665\,D + 4{,}696\,E \\
&\quad + 16{,}752\,H - 8{,}039\,I + 2902{,}27\,M - 459{,}33\,N \\
+370 &= +5{,}917\,A + 17{,}096\,E - 12{,}053\,F - 14{,}707\,G - 4{,}874\,I \\
&\quad - 11{,}963\,K + 30{,}859\,L - 459{,}33\,M + 3385{,}96\,N
\end{aligned}
$$

Hinc eruimus per eliminationem:

$$
\begin{aligned}
A &= -0{,}598 & H &= +0{,}659 \\
B &= -0{,}255 & I &= +1{,}050 \\
C &= -1{,}234 & K &= +0{,}577 \\
D &= +0{,}086 & L &= -1{,}351 \\
E &= -0{,}447 & M &= -0{,}109792 \\
F &= +1{,}351 & N &= +0{,}119681 \\
G &= +0{,}271
\end{aligned}
$$

Jacobi also Invents an Iterative Method

Jacobi (1845): Ueber eine neue Auflösungsart der bei der Methode der kleinsten Quadrate vorkommenden lineären Gleichungen



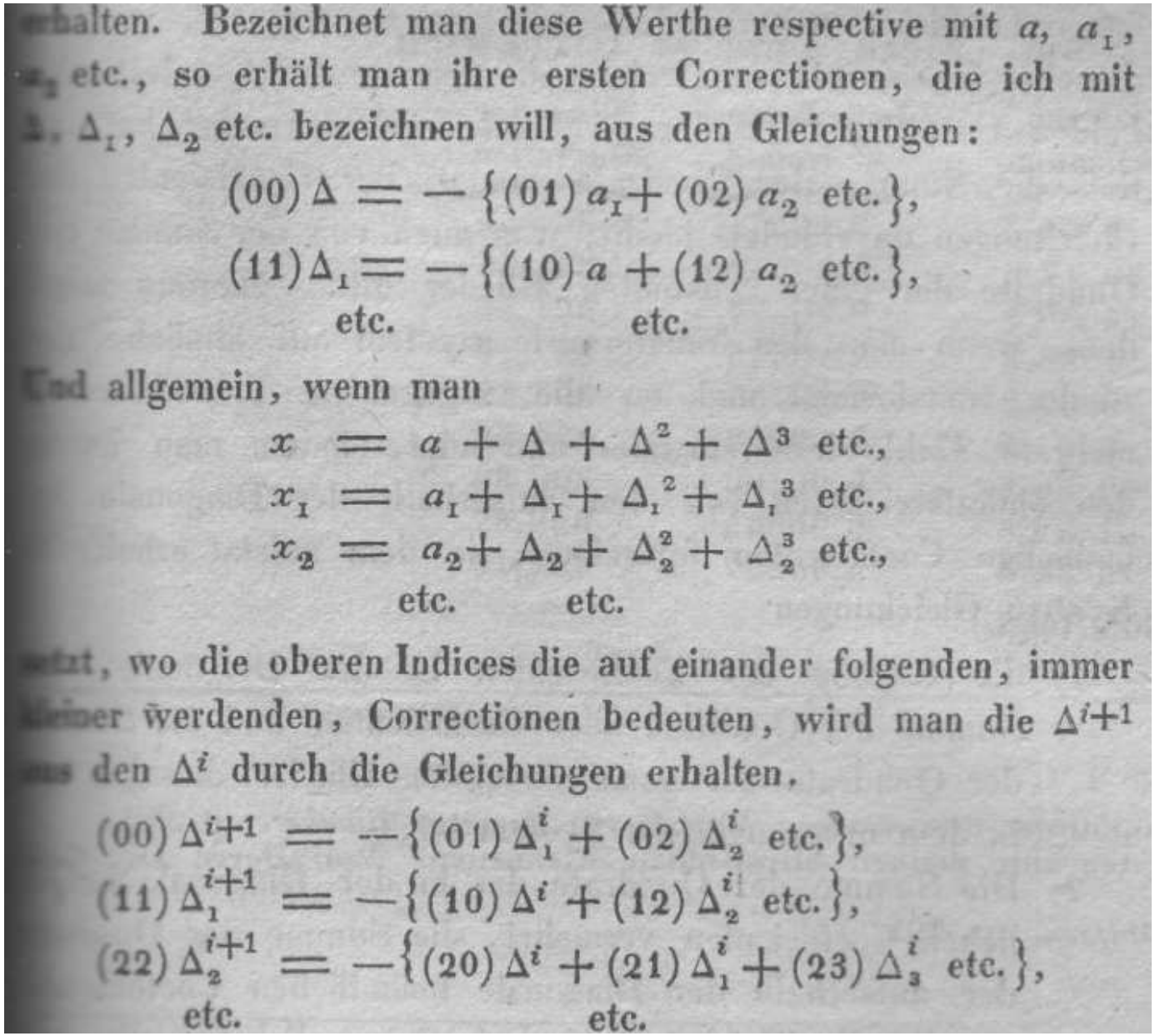
Die Beschwerlichkeit der strengen Auflösung einer gröfseren Zahl lineärer Gleichungen, auf welche in vielen Fällen die Methode der kleinsten Quadrate führt, hat an die Anwendung von Näherungsmethoden denken lassen. Eine solche bietet sich von selber dar, wenn in den verschiedenen Gleichungen immer eine andere Variable mit einem vorzugsweise grofsen Coefficienten multiplicirt ist. Es seien nämlich die Gleichungen:

$$\begin{aligned}
(00)\,x + (01)\,x_1 + (02)\,x_2 \text{ etc.} &= (0m),\\
(10)\,x + (11)\,x_1 + (12)\,x_2 \text{ etc.} &= (1m),\\
(20)\,x + (21)\,x_1 + (22)\,x_2 \text{ etc.} &= (2m),\\
\text{etc.} \quad \text{etc.} \quad \text{etc.},&
\end{aligned}$$

und alle Coefficienten (ik) gegen die in der Diagonale befindlichen (ii) sehr klein, so wird man einen Näherungswerth der unbekannten x, x_1, x_2 etc. aus den Gleichungen:

$$(00)\,x = (0m), \quad (11)\,x_1 = (1m), \quad (22)\,x_2 = (2m), \text{ etc.}$$

Jacobi's Method



erhalten. Bezeichnet man diese Werthe respective mit a, a_1, a_2 etc., so erhält man ihre ersten Correctionen, die ich mit Δ, Δ_1, Δ_2 etc. bezeichnen will, aus den Gleichungen:

$$(00)\,\Delta = -\{(01)\,a_1 + (02)\,a_2 \text{ etc.}\},$$
$$(11)\,\Delta_1 = -\{(10)\,a + (12)\,a_2 \text{ etc.}\},$$
$$\text{etc.} \qquad \text{etc.}$$

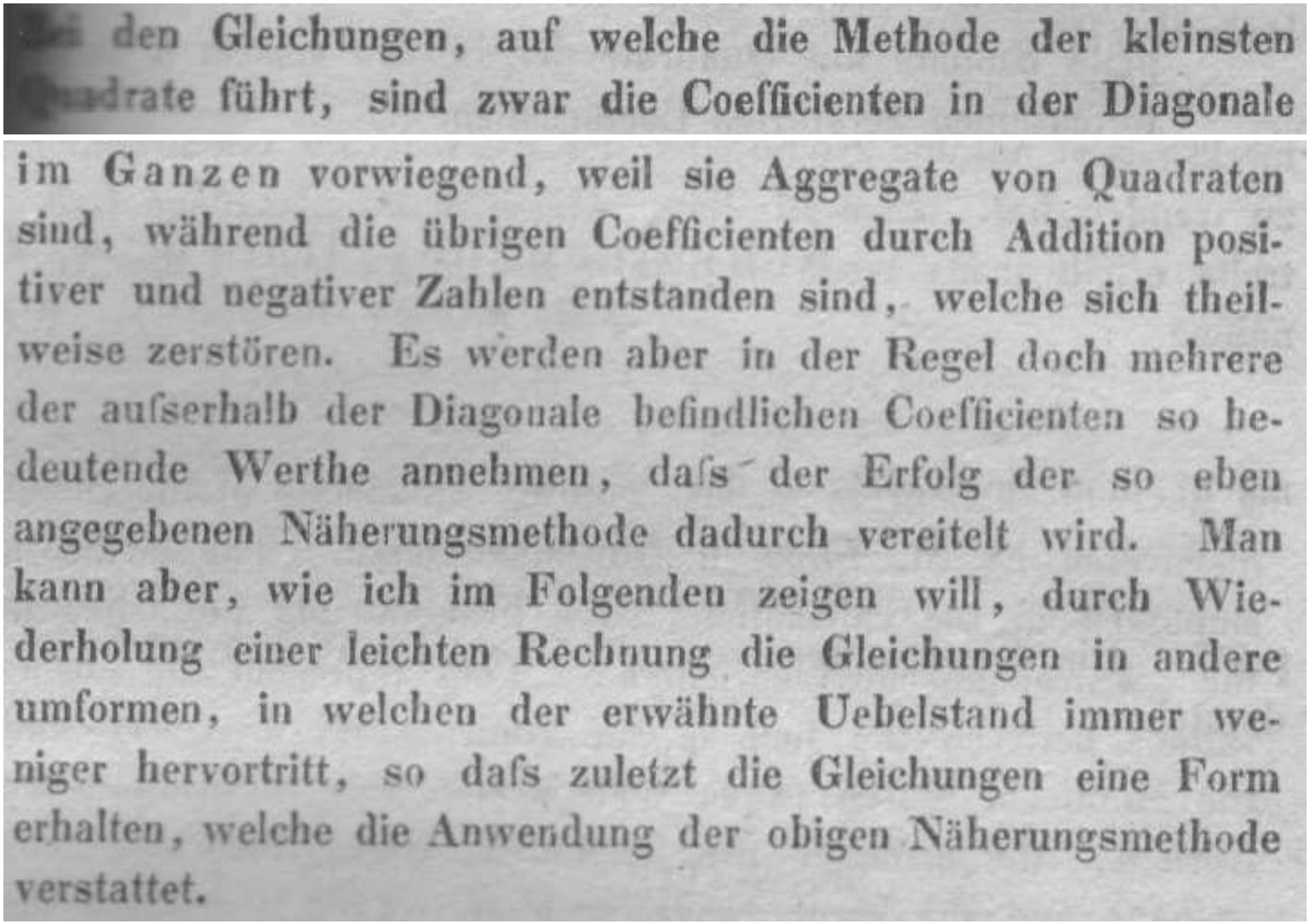
Und allgemein, wenn man

$$x = a + \Delta + \Delta^2 + \Delta^3 \text{ etc.},$$
$$x_1 = a_1 + \Delta_1 + \Delta_1^2 + \Delta_1^3 \text{ etc.},$$
$$x_2 = a_2 + \Delta_2 + \Delta_2^2 + \Delta_2^3 \text{ etc.},$$
$$\text{etc.} \qquad \text{etc.}$$

setzt, wo die oberen Indices die auf einander folgenden, immer kleiner werdenden, Correctionen bedeuten, wird man die Δ^{i+1} aus den Δ^i durch die Gleichungen erhalten.

$$(00)\,\Delta^{i+1} = -\{(01)\,\Delta_1^i + (02)\,\Delta_2^i \text{ etc.}\},$$
$$(11)\,\Delta_1^{i+1} = -\{(10)\,\Delta^i + (12)\,\Delta_2^i \text{ etc.}\},$$
$$(22)\,\Delta_2^{i+1} = -\{(20)\,\Delta^i + (21)\,\Delta_1^i + (23)\,\Delta_3^i \text{ etc.}\},$$
$$\text{etc.} \qquad \text{etc.}$$

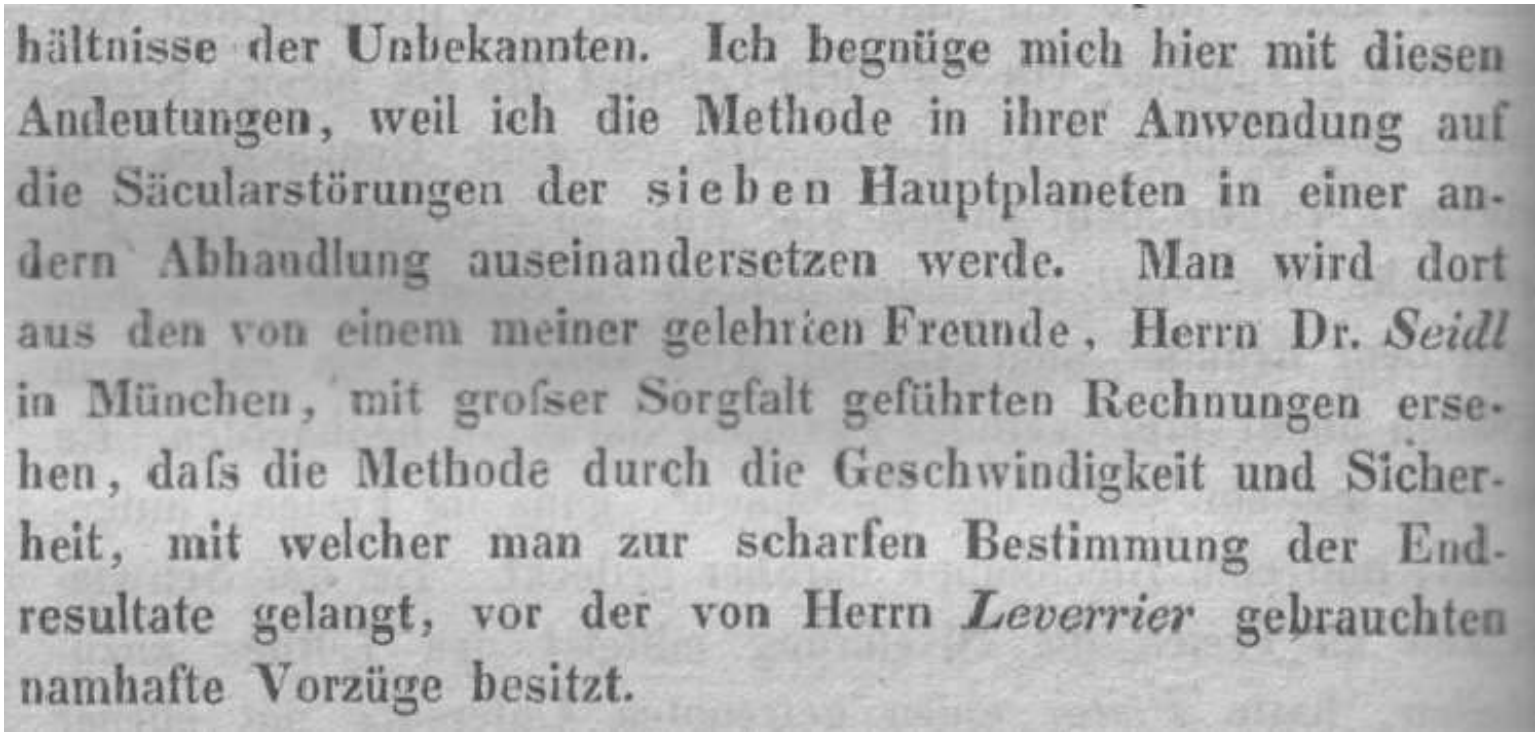
New Idea of Jacobi



den Gleichungen, auf welche die Methode der kleinsten
drate führt, sind zwar die Coefficienten in der Diagonale
im Ganzen vorwiegend, weil sie Aggregate von Quadraten
sind, während die übrigen Coefficienten durch Addition posi-
tiver und negativer Zahlen entstanden sind, welche sich theil-
weise zerstören. Es werden aber in der Regel doch mehrere
der aufserhalb der Diagonale befindlichen Coefficienten so be-
deutende Werthe annehmen, dafs der Erfolg der so eben
angegebenen Näherungsmethode dadurch vereitelt wird. Man
kann aber, wie ich im Folgenden zeigen will, durch Wie-
derholung einer leichten Rechnung die Gleichungen in andere
umformen, in welchen der erwähnte Uebelstand immer we-
niger hervortritt, so dafs zuletzt die Gleichungen eine Form
erhalten, welche die Anwendung der obigen Näherungsmethode
verstattet.

. . . and he invents the Jacobi rotations !

Jacobi Rotations



hältnisse der Unbekannten. Ich begnüge mich hier mit diesen Andeutungen, weil ich die Methode in ihrer Anwendung auf die Säcularstörungen der sieben Hauptplaneten in einer andern Abhandlung auseinandersetzen werde. Man wird dort aus den von einem meiner gelehrten Freunde, Herrn Dr. *Seidl* in München, mit grofser Sorgfalt geführten Rechnungen ersehen, dafs die Methode durch die Geschwindigkeit und Sicherheit, mit welcher man zur scharfen Bestimmung der Endresultate gelangt, vor der von Herrn *Leverrier* gebrauchten namhafte Vorzüge besitzt.

Next, Jacobi takes an example from Gauss' Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium (1809)

Jacobi does Preconditioning

Als ein Beispiel möge hier die Anwendung der Methode auf die in der Theoria motus p. 219 gegebenen Gleichungen dienen. Die ursprünglichen Gleichungen sind

$$\begin{aligned} 27\,p + 6\,q + *\,r - 88 &= 0 \\ 6\,p + 15\,q + r - 70 &= 0 \\ *\,p + q + 54\,r - 107 &= 0. \end{aligned}$$

Schafft man den Coefficienten 6 bei q in der ersten Gleichung fort, so wird $\alpha = 22^\circ 30'$

$$\begin{aligned} p &= 0{,}92390\,y + 0{,}38268\,y' \\ q &= 0{,}38268\,y - 0{,}92390\,y' \end{aligned}$$

und die neuen Gleichungen werden

$$\begin{aligned} 29{,}4853\,y + * \;\; y' + 0{,}38268\,r - 108{,}0901 &= 0 \\ * \;\; y + 12{,}5147\,y' - 0{,}92390\,r + 30{,}9967 &= 0 \\ 0{,}38268\,y - 0{,}92390\,y' + 54{,}r - 107 &= 0 \end{aligned}$$

After preconditioning, it takes only three Jacobi iterations to obtain three accurate digits!

Ludwig Seidel

Seidel (1874): Ueber ein Verfahren, die Gleichungen, auf welche die Methode der kleinsten Quadrate führt, sowie lineäre Gleichungen ueberhaupt, durch successive Annäherung aufzulösen

So einfach nun, ihrer mathematischen Natur nach, die Aufgabe ist, eine beliebige Anzahl unbekannter Grössen aus gleich vielen linearen Gleichungen zu berechnen, so mühsam wird ihre numerische Durchführung, wenn die Zahl der Unbekannten beträchtlich gross wird,

der anzuschliessen. Ich weiss nicht, ob ein Complex von mehr als einigen siebzig Unbekannten je einheitlich berechnet worden ist. Die Zahl 70 ist erreicht in dem Netze der ostpreussischen Gradmessung[1]) (und zwar in einem Falle, wo zwischen den Unbekannten noch 31 streng zu erfüllende Bedingungsgleichungen bestehen, welcher Umstand aber nach der gewöhnlichen Art der Behandlung die Sache nur erschwert), und mit 72 Unbekannten habe ich zu thun gehabt bei Berechnung der wahrscheinlichsten Werthe für die Logarithmen der Helligkeiten der Sterne, welche in mein photometrisches Netz gezogen waren[2]) — Die gebräuch-

Seidel was Jacobi's Student

Seidel talks about Jacobi, and how he had the honor to calculate for him:

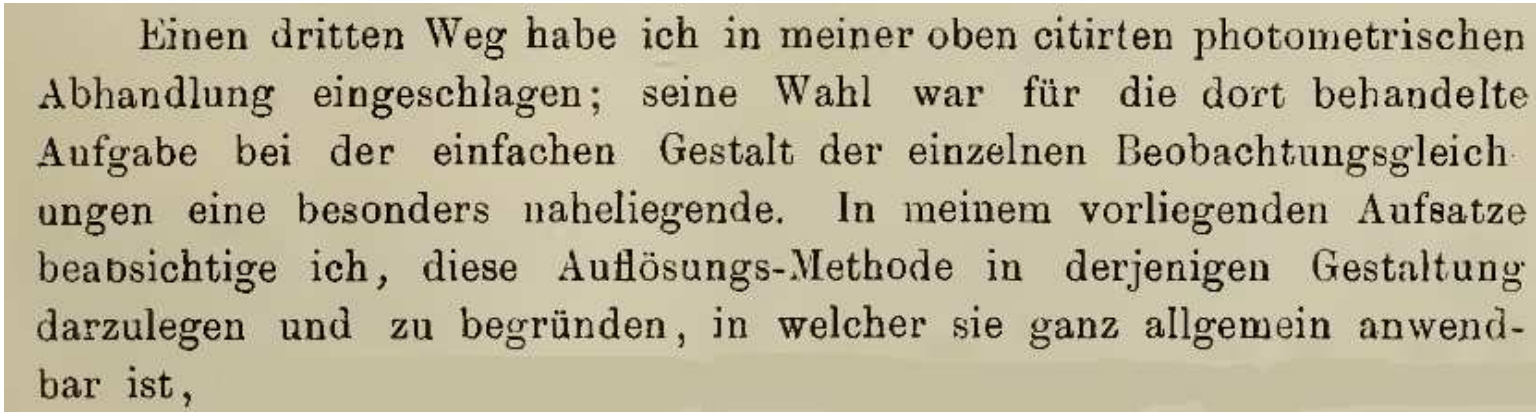
Ein anderes Verfahren hat Jacobi erdacht und auf die 7 Gleichungen angewandt, welche zur Berechnung eines Theiles der Säcularstörungen im Planetensystem nach Laplace von Leverrier aufgestellt waren[1]); ich habe noch als Studirender die Ehre gehabt, für ihn dazu die numerischen Rechnungen auszuführen.

but he also did not think much about the preconditioning idea of Jacobi:

So sinnreich übrigens diese Methode den speciellen Schwierigkeiten des Falles angepasst ist, für welchem Jacobi ihre Anwendung veranlasste und in welchem die diagonalen Coefficienten selbst noch lineäre Funktionen einer supernumerären Unbekannten sind, so scheint sie mir für den gewöhnlich vorkommenden einfacheren Fall doch keineswegs vortheilhafter zu sein, als die allgemein angewandte; auch bezweifle ich, ob sie in irgend einem weiteren Falle bisher in Anwendung gebracht worden ist.

Seidel's "New Method"

Seidel then proposes a third method, his method:



Einen dritten Weg habe ich in meiner oben citirten photometrischen Abhandlung eingeschlagen; seine Wahl war für die dort behandelte Aufgabe bei der einfachen Gestalt der einzelnen Beobachtungsgleichungen eine besonders naheliegende. In meinem vorliegenden Aufsatze beabsichtige ich, diese Auflösungs-Methode in derjenigen Gestaltung darzulegen und zu begründen, in welcher sie ganz allgemein anwendbar ist,

- Seidel describes in what follows Gauss' method
- He gives a convergence prove for the case of the normal equations
- He emphasizes that any system can be written as a system of normal equations
- He notes that the unknowns could also be processed cyclically
- and ...

Seidel Does Parallel Computing!

Verbindung gebracht werden sollen; d. h. man nehme an, dass das Beobachtungsmaterial (niedergelegt in Gleichungen der Form A), welches sehr viele Unbekannte enthalten mag, sich unter zwei Rechner so vertheilen lässt, dass die Gleichungen, welche gewisse Unbekannte enthalten, ausschliesslich dem ersten, diejenigen mit anderen Unbekannten ausschliesslich dem zweiten zugewiesen werden, und dass nur verhältnissmässig wenig Unbekannte einer dritten oder intermediären Gruppe durch Beobachtungsgleichungen einerseits mit Variablen des ersten und andererseits mit solchen des zweiten Systemes in Verbindung gebracht sind, und so den Zusammenhang beider Systeme vermitteln. Die Normalgleichungen (in der Form E, E') für die Unbekannten dieser verbindenden Art müssen von beiden Rechnern (natürlich übereinstimmend) angesetzt werden: ihre Glieder zur Rechten (auch die Complexe $\xi . \eta \ldots$) werden von selbst in drei Theile zerfallen, die Unbekannten des ersten, des zweiten und des intermediären Systems, resp. deren Correctionen, enthaltend. Ausserdem hat jeder Rechner vor sich die Normalgleichungen für die ihm allein zugetheilten Unbekannten. Indem nun ausgeglichen wird, muss von Zeit zu Zeit an gewissen Abschnitten der Arbeit jeder Rechner dem andern für die Normalgleichungen der gemeinsamen Unbekannten die Zahlenwerthe derjenigen Glieder mittheilen,

Modern Notation

Large sparse linear system

$$A\mathbf{x} = \mathbf{b}, \quad A \in \mathbb{R}^{n\times n}, \quad \mathbf{b} \in \mathbb{R}^n,$$

Split the matrix, $A = M - N$, and iterate

$$M\mathbf{x}_{k+1} = N\mathbf{x}_k + \mathbf{b}, \quad k = 0, 1, 2, \ldots$$

Stationary iterative method: M and N do not depend on k

Conditions on M:

- M is a good approximation of A,
- $M\mathbf{x} = \mathbf{y}$ is easy and cheap to solve.

Standard form: $\mathbf{x}_{k+1} = M^{-1}N\mathbf{x}_k + M^{-1}\mathbf{b}$
Correction form: $\mathbf{x}_{k+1} = \mathbf{x}_k + M^{-1}(\mathbf{b} - A\mathbf{x}_k)$

Residual, Error and the Difference of Iterates

Theorem: Defining the

- error: $\mathbf{e}_k := \mathbf{x} - \mathbf{x}_k$
- residual: $\mathbf{r}_k := \mathbf{b} - A\mathbf{x}_k$
- difference of iterates: $\mathbf{u}_k := \mathbf{x}_{k+1} - \mathbf{x}_k$

we have

$$\begin{aligned} \mathbf{e}_{k+1} &= M^{-1}N\mathbf{e}_k, \\ \mathbf{u}_{k+1} &= M^{-1}N\mathbf{u}_k, \\ \mathbf{r}_{k+1} &= NM^{-1}\mathbf{r}_k, \end{aligned}$$

and difference, error and residual are related by

$$M\mathbf{u}_k = \mathbf{r}_k = A\mathbf{e}_k.$$

Theorem (Nekrasov 1885, Pizzetti 1887) The method converges iff $\rho(M^{-1}N) < 1$.

Convergence Factor and Convergence Rate

$$\mathbf{e}_k = (M^{-1}N)^k \mathbf{e}_0 = G^k \mathbf{e}_0$$

taking norms, we want for the error reduction to tolerance ε

$$\frac{\|\mathbf{e}_k\|}{\|\mathbf{e}_0\|} \leq \|G^k\| = \left(\left\|G^k\right\|^{\frac{1}{k}}\right)^k < \varepsilon \Longrightarrow k > \frac{\ln(\varepsilon)}{\ln\left(\|G^k\|^{\frac{1}{k}}\right)}.$$

Lemma: For any induced matrix norm, we have

$$\lim_{k\to\infty} \|G^k\|^{\frac{1}{k}} = \rho(G).$$

Definition:

- mean convergence factor: $\rho_k(G) = \|G^k\|^{\frac{1}{k}}$
- asymptotic convergence factor: $\rho(G) = \lim_{k\to\infty} \rho_k(G)$
- mean convergence rate
  $R_k(G) = -\ln\left(\|G^k\|^{\frac{1}{k}}\right) = -\ln(\rho_k(G))$
- asymptotic convergence rate $R_\infty(G) = -\ln(\rho(G))$

Jacobi's Method

One step of Jacobi for the system $Ax = b$

```
for i=1:n
  tmp(i)=(b(i)-A(i,[1:i-1 i+1:n])*x([1:i-1 i+1:n]))/A(i,i);
end
x=tmp(:);
```

This corresponds to the splitting $A = M - N$ with

$$M = D, \quad N = -L - U \quad \Longrightarrow \quad D\mathbf{x}_{k+1} = \mathbf{b} - (L + U)\mathbf{x}_k$$

Theorem: If the matrix $A \in \mathbb{R}^{n\times n}$ is strictly diagonally dominant , i.e.

$$|a_{ii}| > \sum_{j\neq i} |a_{ij}| \quad \text{for } i = 1, \ldots, n,$$

then the Jacobi iteration converges.

Gauss-Seidel

One step of Gauss-Seidel for the system $Ax = b$

```
for i=1:n
  x(i)=(b(i)-A(i,[1:i-1 i+1:n])*x([1:i-1 i+1:n]))/A(i,i);
end
```

This corresponds to the splitting

$$M = D + L, \quad N = -U$$

$$\Longrightarrow (D + L)\mathbf{x}_{k+1} = -U\mathbf{x}_k + \mathbf{b}.$$

Remark: Gauss-Seidel is often faster than Jacobi, e.g. for the Laplacian, Gauss-Seidel converges twice as fast as Jacobi

Gauss-Seidel always faster than Jacobi ?

$$A = \begin{pmatrix} -1 & 0 & -1 \\ -1 & 1 & 0 \\ 1 & 2 & -3 \end{pmatrix}$$

The Jacobi iteration matrix

$$G_{\mathrm{J}} = -D^{-1}(L+U) = \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 \end{pmatrix},$$

has eigenvalues $0.37 \pm 0.86i$ and -0.74, $\rho(G_{\mathrm{J}}) = 0.944$.
Gauss-Seidel iteration matrix

$$G_{\mathrm{GS}} = -(D+L)^{-1}U = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & -1 \end{pmatrix},$$

has eigenvalues 0, 0, -1 with $\rho(G_{\mathrm{GS}}) = 1$.

Successive Overrelaxation

One step of Successive Overrelaxation (Young 1950) with parameter ω for the system $Ax = b$

```
for i=1:n
  x(i)=omega*(b(i)-A(i,[1:i-1 i+1:n])*x([1:i-1 i+1:n]))/A(i,i)
       +(1-omega)*x(i);
end
```

SOR can also be derived from Gauss-Seidel

$$(D + L)\mathbf{x} = -U\mathbf{x} + \mathbf{b}.$$

Multiplying by ω and adding on both sides the expression $(1 - \omega)D\mathbf{x}$, we obtain the SOR iteration

$$(D + \omega L)\mathbf{x}_{k+1} = (-\omega U + (1 - \omega)D)\mathbf{x}_k + \omega\mathbf{b}.$$

SOR is therefore based on the splitting

$$A = M - N \text{ with } M = \frac{1}{\omega}D + L \text{ and } N = -U + \left(\frac{1}{\omega} - 1\right) D$$

Optimal Choice of Relaxation

Theorem [Young 1950]: Let A have property A, i.e. there exist diagonal matrices D_1 and D_2 s.t

$$\tilde{A} = P^T A P = \begin{bmatrix} D_1 & F \\ E & D_2 \end{bmatrix} = L + D + U.$$

Let $G_J(\tilde{A})$ be the associated Jacobi iteration matrix. If the eigenvalues $\mu(G_J)$ are real and $\rho(G_J) < 1$, then the optimal SOR parameter ω for $\tilde{A}$ is

$$\omega_{\text{opt}} = \frac{2}{1 + \sqrt{1 - \rho(G_J)^2}}.$$

Example: For the discretized Laplacian, Jacobi and Gauss-Seidel have convergence factors $1 - O(h^2)$, and optimized SOR has a convergence factor $1 - O(h)$.

Convergence Results for SOR

Theorem [Kahan (1958)]: For the SOR iteration matrix

$$G_{\mathrm{SOR}} = (D + \omega L)^{-1}(-\omega U + (1-\omega)D)$$

we have

$$\rho(G_{\mathrm{SOR}}) \geq |\omega - 1|, \ \forall \omega.$$

Theorem [Ostrowski-Reich]: Let $A \in \mathbb{R}^{n \times n}$ be symmetric and invertible, with positive diagonal elements, $D > 0$. Then SOR converges for all $0 < \omega < 2$ if and only if A is positive definite.

Richardson Iteration

From the correction form of the stationary iteration

$$\mathbf{x}_{k+1} = \mathbf{x}_k + M^{-1}\mathbf{r}_k,$$

choosing $M^{-1} = \alpha I$, which corresponds to the splitting $M = \frac{1}{\alpha} I$ and $N = \frac{1}{\alpha} I - A$, leads to Richardson's method

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha(\mathbf{b} - A\mathbf{x}_k) = (I - \alpha A)\mathbf{x}_k + \alpha\mathbf{b}.$$

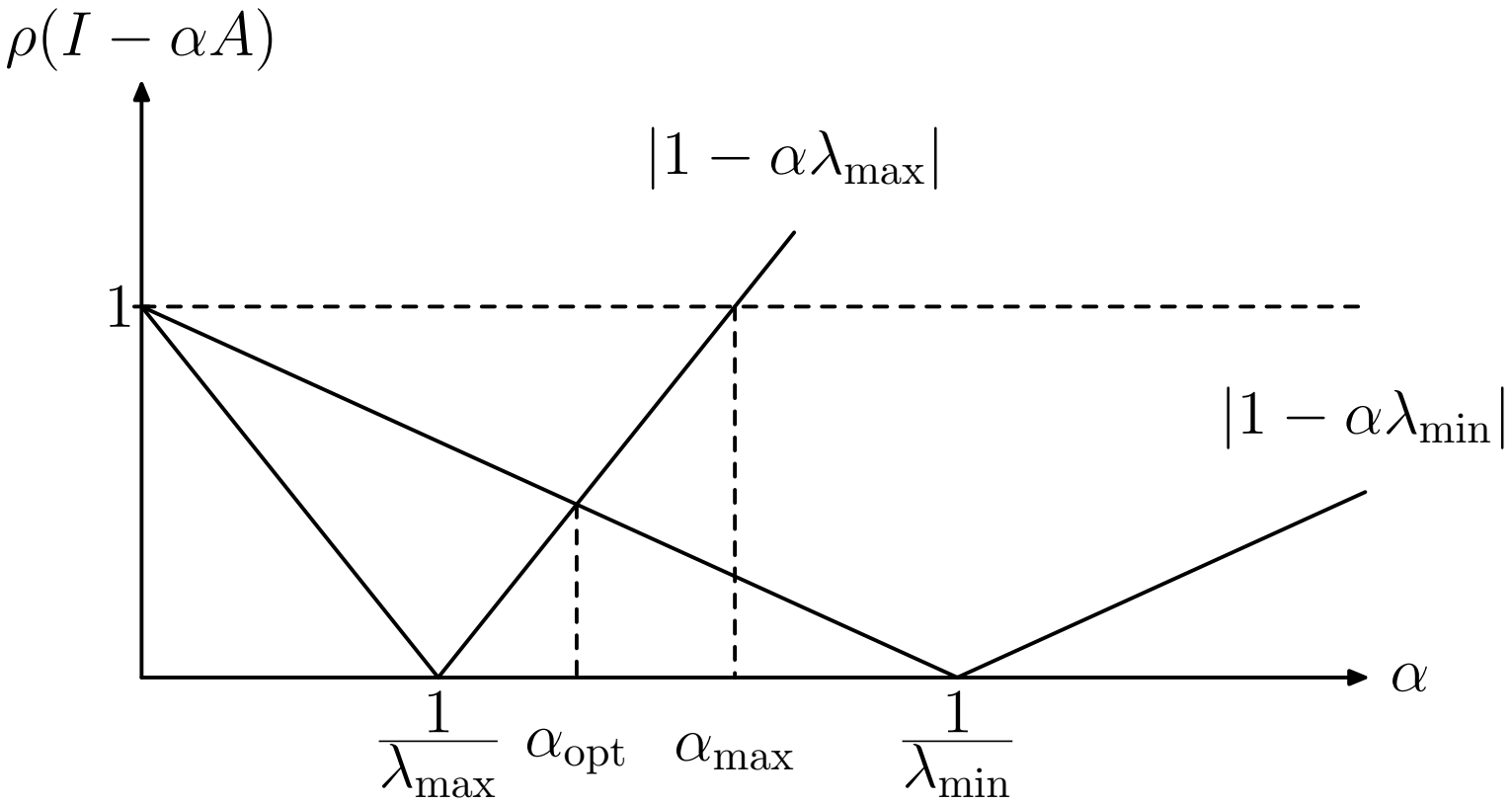
Theorem: For $A \in \mathbb{R}^{n\times n}$ symmetric and positive definite:

a) Richardson converges $\iff$ $0 < \alpha < \frac{2}{\rho(A)}$.

b) Convergence is optimal for $\alpha_{\text{opt}} = \frac{2}{\lambda_{\max}(A)+\lambda_{\min}(A)}$.

c) $\rho(I - \alpha_{\text{opt}} A) = \frac{\kappa(A)-1}{\kappa(A)+1}$, $\kappa(A) := \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$.

Proof of the Theorem



Remark: For the Laplacian, we get the convergence factor $1 - O(h^2)$

But is this really what Richardson proposed ???

Richardson's Influential Paper

Richardson (1910): The Approximate Arithmetical Solution by Finite Differences of Physical Problems involving Differential Equations, with an Application to the stresses in a Masonry Dam

Contains historical section:

§ 2·0. *Historical.*—Step-by-step arithmetical methods of solving ordinary difference equations have long been employed for the calculation of interest and annuities. Recently their application to differential equations has been very greatly improved by the introduction of rules allied to those for approximate quadrature. The papers referred to are :—

RUNGE, "Über die numerische Auflösung von Differentialgleichungen," 'Math. Ann.,' Bd. 46. Leipzig, 1895.

W. F. SHEPPARD, "A Method for Extending the Accuracy of Mathematical Tables," 'Proc. Lond. Math. Soc.,' XXXI.

KARL HEUN, "Neue methode zur approximativen Integration der Differentialgleichungen einer unabhängigen Veränderlichen," 'Zeitschrift Math. u. Phys.,' No. 45, 1900.

WILHELM KUTTA, "Beitrag zur näherungsweisen Integration totaler Differentialgleichen," 'Zeitschrift Math. u. Phys.,' No. 46, 1901.

Further RICHARD GANZ, in a paper "Über die numerische Auflösung von partiellen Differentialgleichungen," 'Zeitschrift Math. u. Phys.,' No. 48, 1903, has extended the methods of RUNGE, HEUN, and KUTTA to partial equations of the type considered in this section.

A Very Rich and Well Written Paper

Focus is the numerical solution of PDEs:

The equations considered in any detail are only a few of the commoner ones occurring in physical mathematics, namely:—LAPLACE's equation $\nabla^2\phi = 0$; the oscillation equations $(\nabla^2+k^2)\,\phi = 0$ and $(\nabla^4-k^4)\,\phi = 0$; and the equation $\nabla^4\phi = 0$. But the methods employed are not limited to these equations.

Explains finite difference approximations:

differencing operator δ and SHEPPARD's* averager μ are defined by

$$\delta f(x) = f(x+\tfrac{1}{2}h) - f(x-\tfrac{1}{2}h) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1),$$

$$\mu f(x) = f(x+\tfrac{1}{2}h) + f(x-\tfrac{1}{2}h) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2),$$

$\frac{\partial\phi}{\partial x}$ will be represented by $\mu\frac{\delta\phi}{\delta x} = \frac{1}{2h}\{(10)-(\bar{1}0)\}$

$\frac{\partial^2\phi}{\partial x^2}$,, ,, ,, $\frac{\delta^2\phi}{\delta x^2} = \frac{1}{h^2}\{(10)-2(00)+(\bar{1}0)\}$

$\frac{\partial^3\phi}{\partial x^3}$,, ,, ,, $\mu\frac{\delta^3\phi}{\delta x^3} = \frac{1}{2h^3}\{(20)-2(10)+2(\bar{1}0)-(\bar{2}0)\}$

$\frac{\partial^4\phi}{\partial x^4}$,, ,, ,, $\frac{\delta^4\phi}{\delta x^4} = \frac{1}{h^4}\{(20)-4(10)+6(00)-4(\bar{1}0)+(\bar{2}0)\}$

$\frac{\partial^2\phi}{\partial x\,\partial y}$,, ,, ,, $\mu^2\frac{\delta^2\phi}{\delta x\,\delta y} = \frac{1}{4h^2}\{(11)+(\bar{1}\bar{1})-(\bar{1}1)-(1\bar{1})\}$

$\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}$,, ,, ,, $\frac{\delta^2\phi}{\delta x^2}+\frac{\delta^2\phi}{\delta y^2} = \triangledown_1{}^2\phi = \frac{1}{h^2}\{(10)+(01)+(\bar{1}0)+(0\bar{1})-4(00)\}.$

Time Stepping Methods for Evolution Problems

One of Richardson's examples: the heat equation

$$u_t = u_{xx} \quad \text{in } \Omega = (-0.5, 0.5) \times (0, 0.005)$$

TABLE I.

	$t = 0.$	0·001.	0·002.	0·003.	0·004.	0·005.	0·005 correct by FOURIER'S method.	Errors.
$x = 0·5$	0·0000	0·0000	0·0000	0·0000	0·0000	0·0000	0·0000	
0·4	1·0000	$a = 0·9090$	0·8356	0·7714	0·7209	0·6729	0·6828	−0·0099
0·3	1·0000	$b = 0·9959$	0·9834	0·9695	0·9492	0·9329	0·9545	−0·0216
0·2	1·0000	$c = 0·9998$	0·9993	0·9968	0·9945	0·9887	0·9980	−0·0093
0·1	1·0000	$d = 1·0000$	1·0000	0·9999	0·9994	0·9990	0·9996	−0·0006
0·0	1·0000	$e = 1·0000$	1·0000	1·0000	1·0000	0·9998	1·0000	−0·0002
−0·1	1·0000	$d = 1·0000$	1·0000	0·9999	0·9994	0·9990	0·9996	−0·0006

In satisfying the equation we must be careful to equate values of $\partial^2\phi/\partial x^2$ and $\partial\phi/\partial t$, which are centered at the same point. This causes a little difficulty at starting. When $t = 0·001$ let the values of ϕ be a, b, c, d, e, as indicated in Table I. Then if the difference equation be satisfied at $t = 0·0005$, it takes the form of 5 simultaneous equations involving a, b, c, d, e. Solving these equations, we find the numbers given in the column $x = 0·001$. Having got over this rather troublesome first step, we can find the rest much more simply by centering all differences on the columns $t = 0·001$, 0·002, 0·003, &c., and deducing each number from the two preceding columns. The

Explains Richardson extrapolation

An excellent illustration is afforded by Lord RAYLEIGH's account of the vibration of a stretched string of beads ('Sound,' vol. I., § 121). He gives the frequency of the fundamental for the same mass per unit length concentrated in various numbers of beads. This is reproduced below in the table. The co-ordinate difference h is inversely as one plus the number of beads, not counting beads at the fixed ends.

Number of free beads + one . .	2	3	4	5	10	20	40	∞
Ratio of frequency to that of continuous string	·9003	·9549	·9745	·9836	·9959	·9990	·9997	unity
Error in representation of continuous string by string of beads	·0997	·0451	·0255	·0164	·0041	·0010	·0003	·0000
Ratio of error to square of co-ordinate difference × a constant	·3988	·4059	·4080	·4091	·4107	·4111	·4112	·4112

The degree of constancy of the last line shows that if we found the frequency for one bead and for three, then extrapolation, on the assumption that the error is proportional to h^2, would give us the frequency for the continuous string to about one part in 1000; which is as near as we could get by twenty beads and no extrapolation.

Solution of Laplace's Equation

TABLE III.

0·5	1	1	1	1	← edge of square.
0	0·5	a	b	c	b
0		0·5	d	e	d
0			0·5	f	0·5
0				centre of 0·5 square.	

↑ edge of square.

TABLE II.

	ψ_1	β_1
ψ_3	ψ_2	β_2
β_4	β_3	

Wall Clock Times of Richardson

The solution of these six simultaneous equations was accomplished in an hour and gave the following results :—

	a.	*b.*	*c.*	*d.*	*e.*	*f.*
By finite differences	0·693	0·772	0·794	0·601	0·6324	0·533
By infinitesimals	0·700	0·777	0·798	0·604	0·6354	0·534
Errors due to finite differences . .	0·007	0·005	0·004	0·003	0·003	0·001

The numbers for infinitesimal differences were obtained from

$$\phi = \frac{4}{\pi} \sum_{m \text{ odd}} (-1)^{\frac{m-1}{2}} \frac{1}{m} \operatorname{sech} \frac{m\pi}{2} \cos mx \cosh mz,$$

"... adding up the series at these six points took 3 hours !"

Richardson also does clamped vibration, on a diamond (comparison with Rayleigh's bound, Ritz is missing!)

Richardson's Iterative Method

§ 3·2. *Successive Approximation to the Integrals.*—Having illustrated the use of simultaneous integral equations, let us pass on to methods which have this property in common: that starting from a table of numbers, correct at the boundary, but otherwise merely as near as one can guess, one proceeds by definite methods to modify this table and thereby to cause it to approach without limit towards the true finite-difference integral.

Conditions of Applicability according to Richardson:

- $\mathcal{D}\phi = 0$, $\mathcal{D}$ a differential operator, together with boundary conditions to make the problem determinate
- Equations must be linear
- Must be such that a certain positive quadratic function is a complete minimum

Let ϕ_u be the correct finite-difference integral. Let ϕ_1 be a function (that is a table of numbers) satisfying the correct boundary-conditions, but arbitrary as to its body values. Next calculate the body-values of ϕ_2 by means of

$$\phi_2 = \phi_1 - \alpha_1^{-1}\,\mathfrak{D}'\phi_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where α_1 is a number to be fixed; and fill in such boundary-values of ϕ_2 as will satisfy the same boundary-conditions as ϕ_u. The succeeding steps are each of the form

$$\phi_{m+1} = \phi_m - \alpha_m^{-1}\,\mathfrak{D}'\phi_m \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Richardson's Analysis

that by the judicious choice of $\alpha_1, \alpha_2, \ldots, \alpha_t$ it is possible to make ϕ_{t+1} nearer to ϕ_u than ϕ_1 was. For since $\mathfrak{D}'$ is linear and $\mathfrak{D}'\phi_u = 0$ we have from (2)

$$\phi_{m+1} - \phi_u = \phi_m - \phi_u - \alpha_m^{-1} \mathfrak{D}' (\phi_m - \phi_u). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3).$$

Now it is shown in the Appendix that $\phi_m - \phi_u$ may be expanded in a series of integrals of

$$(\mathfrak{D}' - \lambda_k^2) \, \mathrm{P}_k = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4).$$

Put $$\phi_1 - \phi_u = \Sigma \mathrm{A}_k \mathrm{P}_k \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5).$$

Then by (4) $$\mathfrak{D}' (\phi_1 - \phi_u) = + \Sigma \mathrm{A}_k \lambda_k^2 \mathrm{P}_k.$$

And therefore by (3) $$\phi_2 - \phi_u = \Sigma \mathrm{A}_k \left(1 - \frac{\lambda_k^2}{\alpha_1}\right) \mathrm{P}_k . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6).$$

Proceeding in the same manner after t operations we arrive at

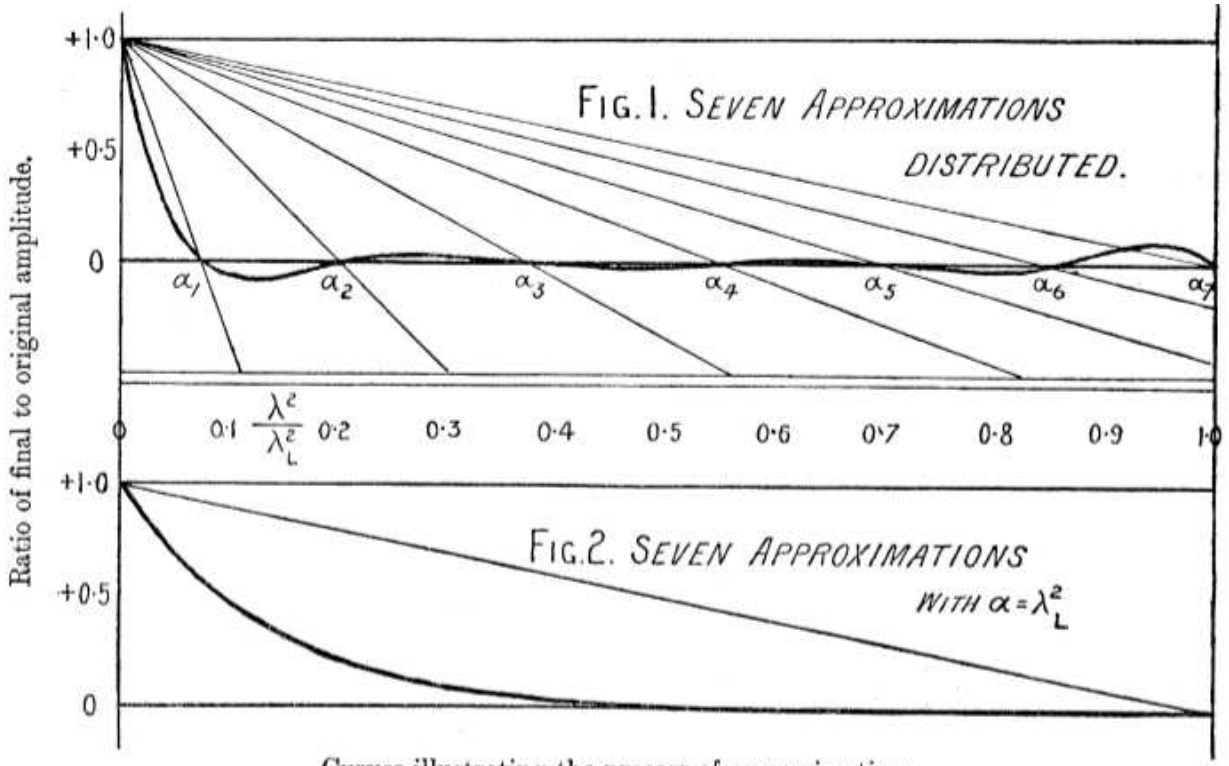
$$\phi_{t+1} - \phi_u = \Sigma \mathrm{A}_k \left(1 - \frac{\lambda_k^2}{\alpha_1}\right)\left(1 - \frac{\lambda_k^2}{\alpha_2}\right) \ldots \left(1 - \frac{\lambda_k^2}{\alpha_t}\right) \mathrm{P}_k \quad . \quad . \quad . \quad . \quad . \quad . \quad (7).$$

A measure of the deviation of two functions from one another which is used in the theory of Least Squares is the sum of the weighted squares of their differences.

Richardson's Minimization

$$\mathrm{SI}\,(\phi_{t+1}-\phi_u)^2 = \Sigma A_k^2 \left[\left(1-\frac{\lambda_k^2}{\alpha_1}\right)\left(1-\frac{\lambda_k^2}{\alpha_2}\right)\times\ldots\times\left(1-\frac{\lambda_k^2}{\alpha_t}\right)\right]^2 \quad . \quad . \quad . \quad (9).$$

Now it has been found that by a judicious choice of $\alpha_1, \alpha_2, \ldots, \alpha_t$, the quantity $\left[\left(1-\frac{\lambda_k^2}{\alpha_1}\right)\left(1-\frac{\lambda_k^2}{\alpha_2}\right)\times\ldots\times\left(1-\frac{\lambda_k^2}{\alpha_t}\right)\right]^2$ may be made small for all possible values of λ_k^2. (Thus fig. 1 shows this done for a set of seven (α)'s. This graph was arrived at by trial.) The error E_{t+1} of ϕ_{t+1} may therefore be made small in comparison with that



Curves illustrating the process of approximation.

Cost per Flop According to Richardson

§ 3·2·5. *Routine of Approximation.—Time and Cost.*—To anyone setting out on a problem I offer the following experience as a guide in forming estimates :—It was found convenient to enter certain stages on a table with large squares, each divided into compartments. Thus for $\frac{\delta^4\phi}{\partial x^4}+2\frac{\delta^4\phi}{\delta x^2\delta y^2}+\frac{\delta^4\phi}{\delta y^4}=0$, one of the squares is shown in the annexed table. All the quantities in it refer to the central point of the square. The intermediate stages are done on rough paper and thrown away. So far I have paid piece rates for the operation $\delta_x^2+\delta_y^2$ of about $\frac{n}{18}$ pence per co-ordinate point, n being the number of digits. The chief trouble to the computers has been the intermixture of plus and minus signs. As to the rate of working, one of the quickest boys averaged 2,000 operations $\delta_x^2+\delta_y^2$ per week, for numbers of three digits, those done wrong being discounted.

TABLE.

ϕ_1	$\nabla^2\phi_1$	$\nabla^4\phi_1$
ϕ_2	$\nabla^2\phi_2$	$\nabla^4\phi_2$
ϕ_3	&c.	

Experience of Richardson:

- $\frac{n}{18}$ pence per coordinate relaxation of Δ_h (n number of digits)
- quickest boys averaged 2000 relaxations of Δ_h per week with 3 digits (incorrect ones discounted)

Real Richardson's Method

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k(\mathbf{b} - A\mathbf{x}_k) \iff \mathbf{r}_{k+1} = (I - \alpha_k A)\mathbf{r}_k$$

Assume A symmetric and positive definite, and define for $\mu \in \mathbb{R}$ the norm

$$\|\mathbf{r}\|^2_{A^{-\mu}} = \mathbf{r}^T A^{-\mu}\mathbf{r}$$

Using Ritz' idea:

$$Q(\alpha_k) := \|\mathbf{r}_{k+1}\|^2_{A^{-\mu}} \longrightarrow \min .$$

Differentiating with respect to α_k,

$$\frac{dQ}{d\alpha_k} = 2\mathbf{r}^T_{k+1}A^{-\mu}\frac{d\mathbf{r}_{k+1}}{d\alpha_k} = 2((I - \alpha_k A)\mathbf{r}_k)^T A^{-\mu}(-A\mathbf{r}_k) = 0$$

Solving for α_k, we obtain

$$\alpha_k = \frac{\mathbf{r}_k^T A^{1-\mu}\mathbf{r}_k}{\mathbf{r}_k^T A^{2-\mu}\mathbf{r}_k}.$$

Useful choices for μ are 0 and 1

Conjugate Residuals

For $\mu = 0$, we obtain

$$\alpha_k = \frac{\mathbf{r}_k^T A \mathbf{r}_k}{\|A\mathbf{r}_k\|^2},$$

and we are minimizing in this case locally

$$\|\mathbf{r}_{k+1}\|_2^2 = \|A\mathbf{e}_{k+1}\|_2^2 = \|\mathbf{e}_{k+1}\|_{A^2}^2 \longrightarrow \min.$$

This algorithm has the property

$$\begin{aligned} \mathbf{r}_{k+1}^T A \mathbf{r}_k &= \mathbf{r}_k^T (I - \alpha_k A)^T A \mathbf{r}_k \\ &= \mathbf{r}_k^T A \mathbf{r}_k - \mathbf{r}_k^T \frac{\mathbf{r}_k^T A \mathbf{r}_k}{\|A\mathbf{r}_k\|^2} A^T A \mathbf{r}_k \;=\; 0 \end{aligned}$$

$\Longrightarrow$ **conjugate residuals algorithm**

Steepest Descent

For $\mu = 1$, we obtain

$$\alpha_k = \frac{\|\mathbf{r}_k\|^2}{\mathbf{r}_k^T A \mathbf{r}_k} \quad \text{and} \quad \|\mathbf{r}_{k+1}\|_{A^{-1}}^2 = \mathbf{r}_{k+1}^T A^{-1} \mathbf{r}_{k+1} \to \min$$

We also minimize $\|\mathbf{e}_{k+1}\|_A^2 = \mathbf{e}_{k+1}^T A \mathbf{e}_{k+1}$, since $A\mathbf{e}_{k+1} = \mathbf{r}_{k+1}$.
We also minimize $Q(\mathbf{x}_k + \alpha \mathbf{r}_k)$ as a function of α, where $Q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A \mathbf{x} - \mathbf{b}^T \mathbf{x}$, since

$$\frac{dQ}{d\alpha} = \mathbf{r}^T A \mathbf{x} + \alpha \mathbf{r}^T A \mathbf{r} - \mathbf{b}^T \mathbf{r} = 0,$$

leads to the same α.

$$\nabla Q = A\mathbf{x} - \mathbf{b} = -\mathbf{r}$$

$\Longrightarrow$ **method of steepest descent**

Theorem

Steepest descent converges for symmetric positive definite problems.

Best Tactic versus Best Strategy

Choosing α_k such that

$$\|\mathbf{r}_{k+1}\|^2_{A^{-\mu}} \longrightarrow \min$$

is the *best tactic* but maybe not the *best strategy*.
Residual recurrence in Richardson's method:

$$\mathbf{r}_{k+1} = (I - \alpha_k A)\mathbf{r}_k,$$

and since $\mathbf{r}_k = A\mathbf{e}_k$

$$\mathbf{e}_{k+1} = (I - \alpha_k A)\mathbf{e}_k.$$

Thus the error after k steps is

$$\mathbf{e}_k = (I - \alpha_{k-1}A)(I - \alpha_{k-2}A)\cdots(I - \alpha_0 A)\mathbf{e}_0 = P_k(A)\mathbf{e}_0.$$

Definition

The polynomial P_k, which satisfies $P_k(0) = 1$, is called *residual polynomial*.

Optimal Strategy

Should choose α_k such that the global error is minimized!

Theorem

*If $\alpha_k = \frac{1}{\lambda_k}$, λ_k eigenvalues of A, then convergence is reached in at most n steps (***Proof:** *Cayley-Hamilton)*

Approximation: since

$$\|\mathbf{e}_k\| = \|P_k(A)\mathbf{e}_0\| \leq \|P_k(A)\|\|\mathbf{e}_0\|,$$

can try to minimize $\|P_k(A)\|$. If A is diagonalizable, $A = Q\Lambda Q^{-1}$

$$\|P_k(A)\| = \|QP_k(\Lambda)Q^{-1}\| \leq \|Q\|\|Q^{-1}\|\|P_k(\Lambda)\| = \kappa(Q)\|P_k(\Lambda)\|.$$

For symmetric and positive definite matrices, an interval

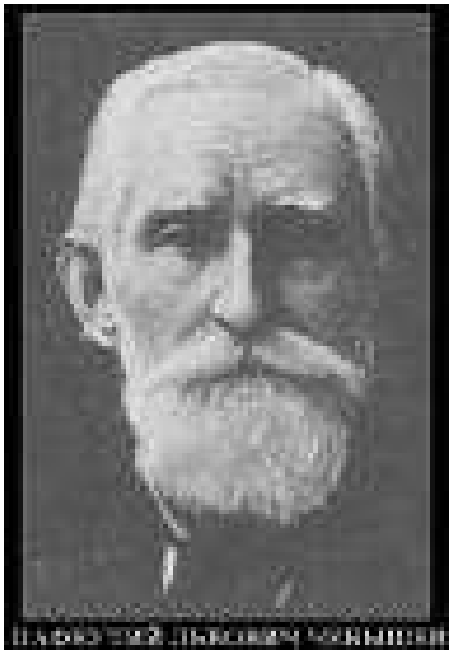
$$0 < a \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq b$$

may be known, and hence

$$\max_i |P_k(\lambda_i)| \leq \max_{a \leq x \leq b} |P_k(x)|.$$

Best Approximation Problems

Chebyshev (1854): Théorie des mécanismes connus sous le nom de parallélogrammes.



Soit $f(x)$ une fonction donnée, U un polynome du degré n avec des coefficients arbitraires. Si l'on choisit ces coefficients de manière à ce que la différence $f(x) - U$, depuis $x = a - h$, jusque à $x = a + h$, reste dans les limites les plus rapprochées de 0, la différence $f(x) - U$ jouira, comme on le sait, de cette propriété:
Parmi les valeurs les plus grandes et les plus petites de la différence $f(x) - U$ entre les limites $x = a - h$, $x = a + h$, on trouve au moins $n+2$ fois la même valeur numérique.

De la Vallée Poussin (1910): Existence, Uniqueness and Equioscillation.

$$\min_{p \in P_n} \max_{x \in K} |f(x) - p(x)|$$

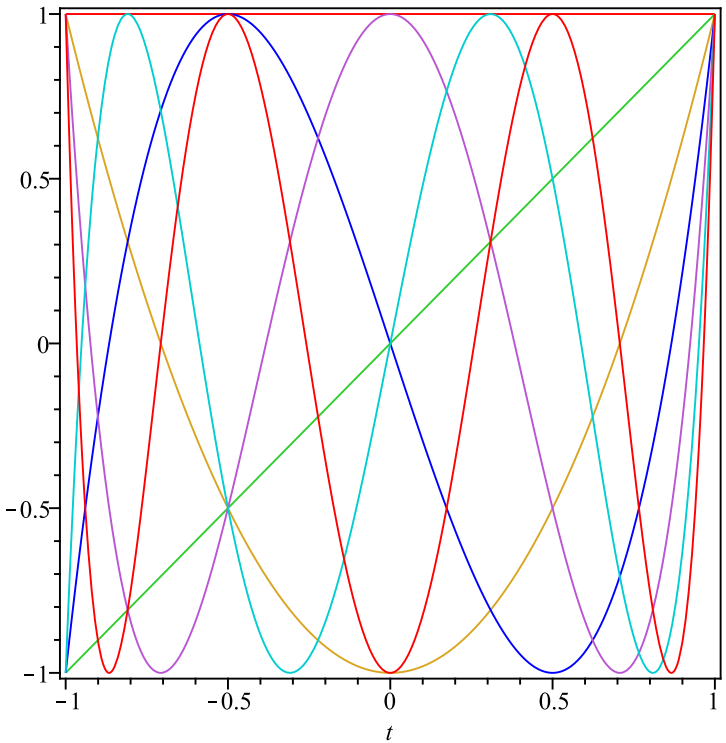
Chebyshev Polynomials

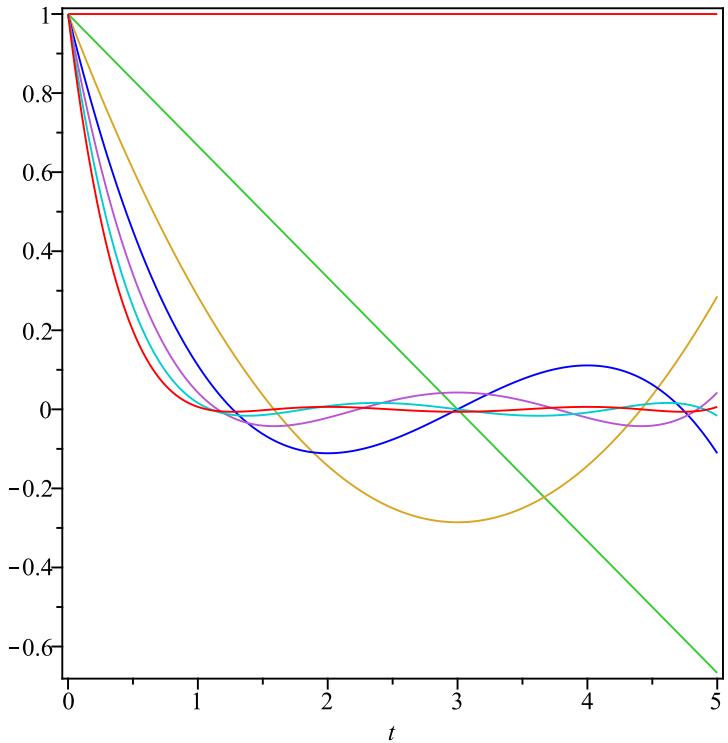
Given a degree k, we want to determine a polynomial P_k with $P_k(0) = 1$ such that

$$\max_{a \le x \le b} |P_k(x)| \longrightarrow \min .$$

Definition (Chebyshev Polynomials)

$$C_k(t) := \cos(k \arccos t), \quad -1 \le t \le 1, \quad k = 0, 1, \ldots$$





Chebyshev Semi-Iterative Method

Normalized Chebyshev polynomial is

$$Q_k(x) = \frac{C_k\left(-1+2\frac{x-a}{b-a}\right)}{C_k\left(\frac{a+b}{a-b}\right)}$$

and can be bounded by

$$\frac{1}{\left|C_k\left(\frac{a+b}{a-b}\right)\right|} \leq 2\left|\frac{\sqrt{\frac{b}{a}}-1}{\sqrt{\frac{b}{a}}+1}\right|^k$$

and we therefore get for this method

$$\|\mathbf{e}_k\| \leq 2\left(\frac{\sqrt{\kappa(A)}-1}{\sqrt{\kappa(A)}+1}\right)^k \|\mathbf{e}_0\|.$$

Richardson's dream comes true in Gene Golub's PhD-thesis (1959): the Chebyshev Semi-Iterative Method!

Extrapolation Methods

Claude Brezinski (1977): Accélération de la Convergence en Analyse Numérique (Springer)
Avram Sidi and Jacob Bridger (1988): Convergence and stability analyses for some vector extrapolation methods in the presence of defective iteration matrices
Avram Sidi (1991): Efficient implementation of minimal polynomial and reduced rank extrapolation methods

For $A\mathbf{x} = \mathbf{b}$, $A = M - N$, we consider the iteration

$$M\mathbf{x}_{i+1} = N\mathbf{x}_i + \mathbf{b}$$

Idea: find γ_i for $i = 0, 1, \ldots, k$, with $\sum_{i=0}^{k} \gamma_i = 1$ such that

$$\mathbf{y}_k := \sum_{i=0}^{k} \gamma_i \mathbf{x}_i$$

is a much better approximation than $\mathbf{x}_k$.

Finding a Good Combination

$$\mathbf{x}_i = \mathbf{x} + \mathbf{e}_i \quad \Longrightarrow \quad \underbrace{\sum_{i=0}^{k} \gamma_i \mathbf{x}_i}_{\mathbf{y}_k} = \mathbf{x} \underbrace{\sum_{i=0}^{k} \gamma_i}_{1} + \sum_{i=0}^{k} \gamma_i \mathbf{e}_i,$$

and therefore

$$\mathbf{y}_k = \mathbf{x} + \sum_{i=0}^{k} \gamma_i \mathbf{e}_i.$$

For $\mathbf{y}_k$ to be a good approximation, we would like to have

$$\sum_{i=0}^{k} \gamma_i \mathbf{e}_i \approx 0,$$

and with the error recursion $\mathbf{e}_i = M^{-1} N \mathbf{e}_{i-1} =: G \mathbf{e}_{i-1}$,

$$\sum_{i=0}^{k} \gamma_i G^i \mathbf{e}_0 = P_k(G) \mathbf{e}_0 \approx 0.$$

An Important Lemma

Lemma

If $P_k(G)\mathbf{e}_0 = 0$, *then for* $m \geq 0$ *also* $P_k(G)\mathbf{e}_m = 0$ *and* $P_k(G)\mathbf{u}_m = 0$, *where* $\mathbf{u}_m := \mathbf{x}_{m+1} - \mathbf{x}_m$.

Proof.

From the recurrence relation, we have

$$\mathbf{u}_m = G\mathbf{u}_{m-1}, \quad \text{and} \quad \mathbf{u}_m = (I - G)\mathbf{e}_m,$$

and since polynomials in G commute, we obtain from $P_k(G)\mathbf{e}_0 = 0$ that

$$0 = G^m P_k(G)\mathbf{e}_0 = P_k(G)G^m\mathbf{e}_0 = P_k(G)\mathbf{e}_m,$$

and similarly

$$0 = (I - G)P_k(G)\mathbf{e}_m = P_k(G)(I - G)\mathbf{e}_m = P_k(G)\mathbf{u}_m.$$

An Extrapolation Method

Idea: try to approximate $P_k(G)\mathbf{u}_0 \approx 0$. Written with the coefficients of P_k:

$$P_k(G)\mathbf{u}_0 = \gamma_0\mathbf{u}_0 + \gamma_1 \underbrace{G\mathbf{u}_0}_{\mathbf{u}_1} + \cdots + \gamma_k \underbrace{G^k\mathbf{u}_0}_{\mathbf{u}_k} = [\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_k]\boldsymbol{\gamma}.$$

Introducing the matrix $U_k \in \mathbb{R}^{n\times(k+1)}$ defined by

$$U_k := [\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_k],$$

we would like to determine the coefficients γ_i in the vector $\boldsymbol{\gamma} \in \mathbb{R}^{k+1}$ such that

$$U_k\boldsymbol{\gamma} \approx 0, \quad \text{subject to the constraint} \quad \sum_{i=0}^{k} \gamma_i = 1.$$

Solvable Optimization Problem with Constraint

Removing the Constraint

In order to remove the constraint, we parametrize the γ_i by k parameters ξ_i,

$$\gamma_0 = 1 - \xi_0, \ \gamma_1 = \xi_0 - \xi_1, \ldots, \gamma_i = \xi_{i-1} - \xi_i, \ldots, \gamma_k = \xi_{k-1},$$

which leads to $\boldsymbol{\gamma} = S\boldsymbol{\xi} + \mathbf{e}_1$, $\boldsymbol{\xi} \in \mathbb{R}^k$, with the matrix

$$S = \begin{bmatrix} -1 & & & \\ 1 & -1 & & \\ & 1 & \ddots & \\ & & \ddots & -1 \\ & & & 1 \end{bmatrix}, \quad S \in \mathbb{R}^{(k+1)\times k}.$$

With this new parametrization, we obtain

$$U_k\boldsymbol{\gamma} = U_k S\boldsymbol{\xi} + U_k\mathbf{e}_1 \approx 0.$$

Extrapolation Methods Continued

The matrix $U_k S$ consists of the columns

$$U_k S = [-\mathbf{u}_0 + \mathbf{u}_1, -\mathbf{u}_1 + \mathbf{u}_2, \ldots, -\mathbf{u}_{k-1} + \mathbf{u}_k].$$

We set $\mathbf{w}_j = \mathbf{u}_{j+1} - \mathbf{u}_j$ and define

$$W_k := [\mathbf{w}_0, \ldots, \mathbf{w}_k] \in \mathbb{R}^{n \times (k+1)}.$$

and thus obtain

$$W_{k-1}\boldsymbol{\xi} \approx -\mathbf{u}_0,$$

$$\mathbf{y}_k = X_k \boldsymbol{\gamma} = X_k S \boldsymbol{\xi} + X_k \mathbf{e}_1 = U_{k-1}\boldsymbol{\xi} + \mathbf{x}_0$$

Solvable optimization problem without constraint

Krylov Spaces

Definition (Krylov Space)

Let $A \in \mathbb{R}^{n\times n}$ and $\mathbf{r} \in \mathbb{R}^n$. The associated *Krylov space* of dimension k is

$$\mathcal{K}_k(A, \mathbf{r}) = \text{span}\{\mathbf{r}, A\mathbf{r}, A^2\mathbf{r}, \ldots, A^{k-1}\mathbf{r}\}.$$

Since $\mathbf{y}_k = \mathbf{x}_0 + U_{k-1}\boldsymbol{\xi}$, it follows that

$$\mathbf{y}_k - \mathbf{x}_0 = [\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{k-1}]\boldsymbol{\xi} = [\mathbf{u}_0, G\mathbf{u}_0, \ldots, G^{k-1}\mathbf{u}_0]\boldsymbol{\xi},$$

and therefore this difference lies in a Krylov space,

$$\mathbf{y}_k - \mathbf{x}_0 \in \mathcal{K}_k(G, \mathbf{u}_0).$$

If we choose $\mathbf{x}_0 = 0$, we have $\mathbf{x}_1 = G\mathbf{x}_0 + \mathbf{d} = \mathbf{d}$, and thus $\mathbf{u}_0 = \mathbf{x}_1 - \mathbf{x}_0 = \mathbf{d}$, which implies

$$\mathbf{y}_k \in \mathcal{K}_k(G, \mathbf{d}) = \mathcal{K}_k(M^{-1}N, M^{-1}\mathbf{b}) = \mathcal{K}_k(M^{-1}A, M^{-1}\mathbf{b}).$$

Minimal Polynomial Extrapolation (MPE)

In order to solve approximately

$$U_k \gamma \approx 0 \quad \text{s.t.} \sum_{j=0}^{k} \gamma_j = 1,$$

MPE fixes the last coefficient to 1 an solves

$$U_{k-1} \mathbf{c} \approx -\mathbf{u}_k$$

using least squares (γ are $\mathbf{c}$ normalized).

Theorem (MPE $\Longleftrightarrow$ FOM)

With γ from MPE, the preconditioned residual

$$\mathbf{r}_k = \mathbf{d} - (I - G)\mathbf{y}_k = M^{-1}(\mathbf{b} - A\mathbf{y}_k)$$

satisfies $\mathbf{r}_k \perp \mathcal{K}_k(G, \mathbf{u}_0)$, and MPE is equivalent to FOM applied to the preconditioned system $M^{-1}A\mathbf{x} = M^{-1}\mathbf{b}$.

Topological ε-Algorithm (TEA)

In order to solve approximately

$$U_k \gamma \approx 0 \quad \text{s.t.} \quad \sum_{j=0}^{k} \gamma_j = 1.$$

one can also use a Galerkin approach. Let $Q_{k-1} = [\mathbf{q}, G^T\mathbf{q}, \ldots, (G^T)^{k-1}\mathbf{q}] \in \mathbb{R}^{n\times k}$ and solve

$$Q_{k-1}^T U_k \gamma = 0, \qquad \sum_{j=0}^{k} \gamma_j = 1.$$

Theorem (Equivalence of TEA and NSL)

With $A = M - N$, for any given starting vector $\mathbf{v}_0$, applying non-symmetric Lanczos (NSL) to the preconditioned system $M^{-1}A\mathbf{v} = M^{-1}\mathbf{b}$, or applying TEA to the stationary iterative method $M\mathbf{v}_{k+1} = N\mathbf{v}_k + \mathbf{b}$ with $\mathbf{q} := \mathbf{r}_0 = M^{-1}\mathbf{b} - M^{-1}A\mathbf{v}_0$ leads to identical iterates.

Preconditioning Arrives Naturally

Every stationary iteration (like multigrid, domain decomposition, etc.) can be written in the form of a stationary iterative method

$$M\mathbf{x}_{k+1} = N\mathbf{x}_k + \mathbf{b} \iff \mathbf{x}_{k+1} = \mathbf{x}_k + M^{-1}(\mathbf{b} - A\mathbf{x}_k)$$

and can therefore serve as a preconditioner for a Krylov method.

For example the restricted additive Schwarz method:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \sum_{i=1}^{I} \tilde{R}_i^T A_i^{-1} R_i(\mathbf{b} - A\mathbf{x}_k)$$

The matrix M is called a preconditioner, and the converged solution satisfies the preconditioned system

$$M^{-1}A\mathbf{x} = M^{-1}\mathbf{b}$$

Krylov Methods Were not Invented Like This

Stiefel and Rosser 1951: Presentations at a Symposium at the National Bureau of Standards (UCLA)

Hestenes 1951: Iterative methods for solving linear equations

Stiefel 1952: Über einige Methoden der Relaxationsrechnung

Hestenes and Stiefel 1952: Methods of Conjugate Gradients for Solving Linear Systems

"An iterative algorithm is given for solving a system $Ax = k$ of n linear equations in n unknowns. The solution is given in n steps."

General Idea of Relaxation (following Stiefel)

If A is symmetric and positive definite, we have

$$A\mathbf{x} = \mathbf{b} \quad \Longleftrightarrow \quad F(\mathbf{x}) := \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{b}^T\mathbf{x} \longrightarrow \min$$

To solve the minimization problem, a natural relaxation procedure is

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \alpha\mathbf{p}_n$$

where $\mathbf{p}_n$ is a search direction and α is the distance to go along this direction.

Example: The Jacobi method

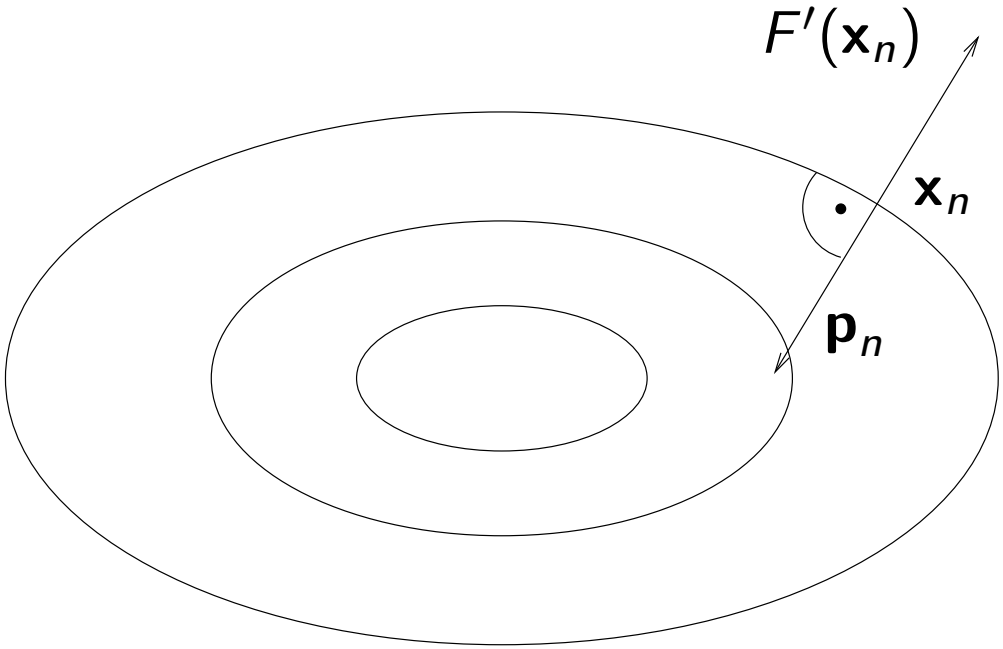
$$\mathbf{x}_{n+1} = \mathbf{x}_n + D^{-1}(\mathbf{b} - A\mathbf{x}_n)$$

for the five point finite difference Laplacian uses

$$\mathbf{p}_n := (\mathbf{b} - A\mathbf{x}_n) \quad \alpha = \frac{1}{4}.$$

Is the Jacobi Choice a Good One ?

The direction of Jacobi is $\mathbf{p} = (\mathbf{b} - A\mathbf{x}_n) = -F'(\mathbf{x}_n)$, and thus Jacobi goes into the direction of fastest decrease of F at $\mathbf{x}^n$:

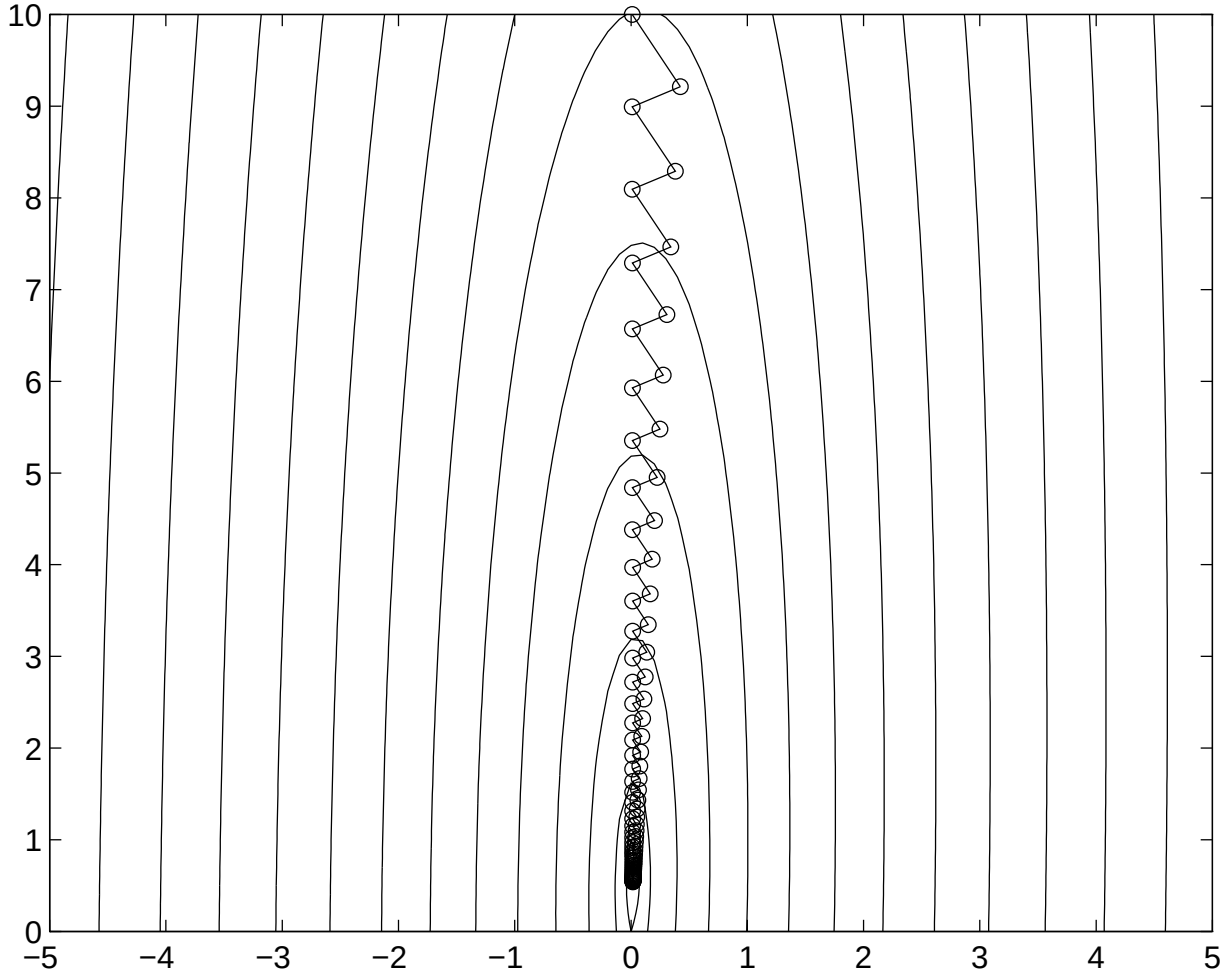


Hence the direction is a good choice, but the distance $\alpha = \frac{1}{4}$ might not be good.

"Ritzscher Gedanke (Stiefel)": Use α to minimize F along the direction $\mathbf{p}$, hence $\alpha = \alpha_n$:

$$\Longrightarrow \textbf{Method of Steepest Descent}.$$

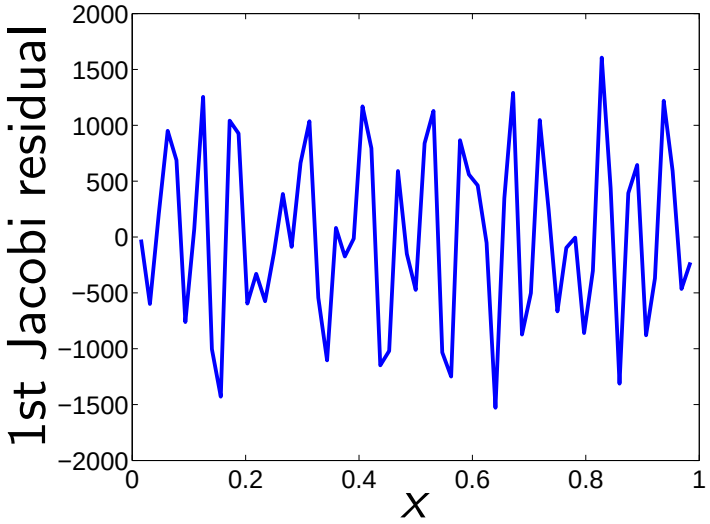
Problems of Steepest Descent: Cage Syndrome

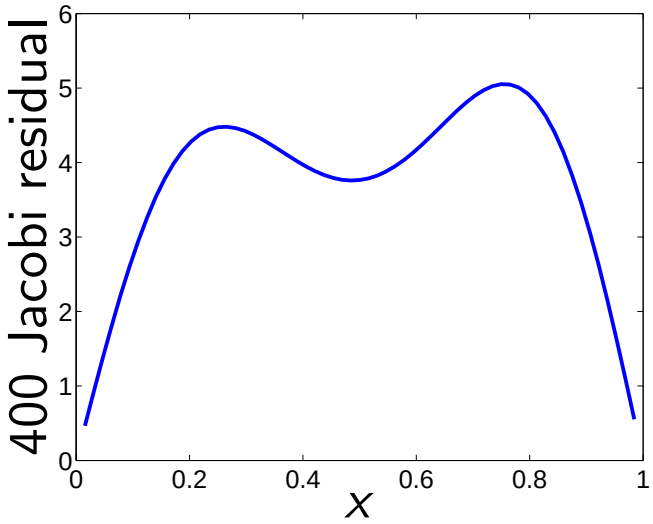


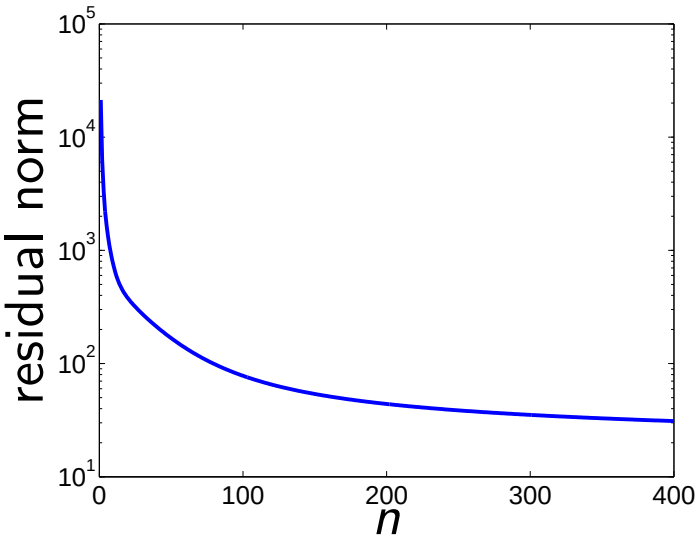
Stiefel 1952: "Das Auftreten von Käfigen ist eine allgemeine Erscheinung bei Relaxationsverfahren und sehr unerwünscht. Es bewirkt, dass eine Relaxation am Anfang flott vorwärts geht, aber dann immer weniger ausgiebig wird..."

General Cage Syndrome for Relaxation Methods

"... Es ist leicht festzustellen, wann man bei diesem Einzelschrittverfahren [Jacobi] des Dirichlet Problems in einem Käfig sitzt: Dies ist dann der Fall, wenn die Residuen in den inneren Punkten alle dieselben Vorzeichen haben."







"... so dass der positive Residualberg mit den Löffel statt mit einer Baggermaschine abgetragen wird !"

Remedies Proposed By Stiefel

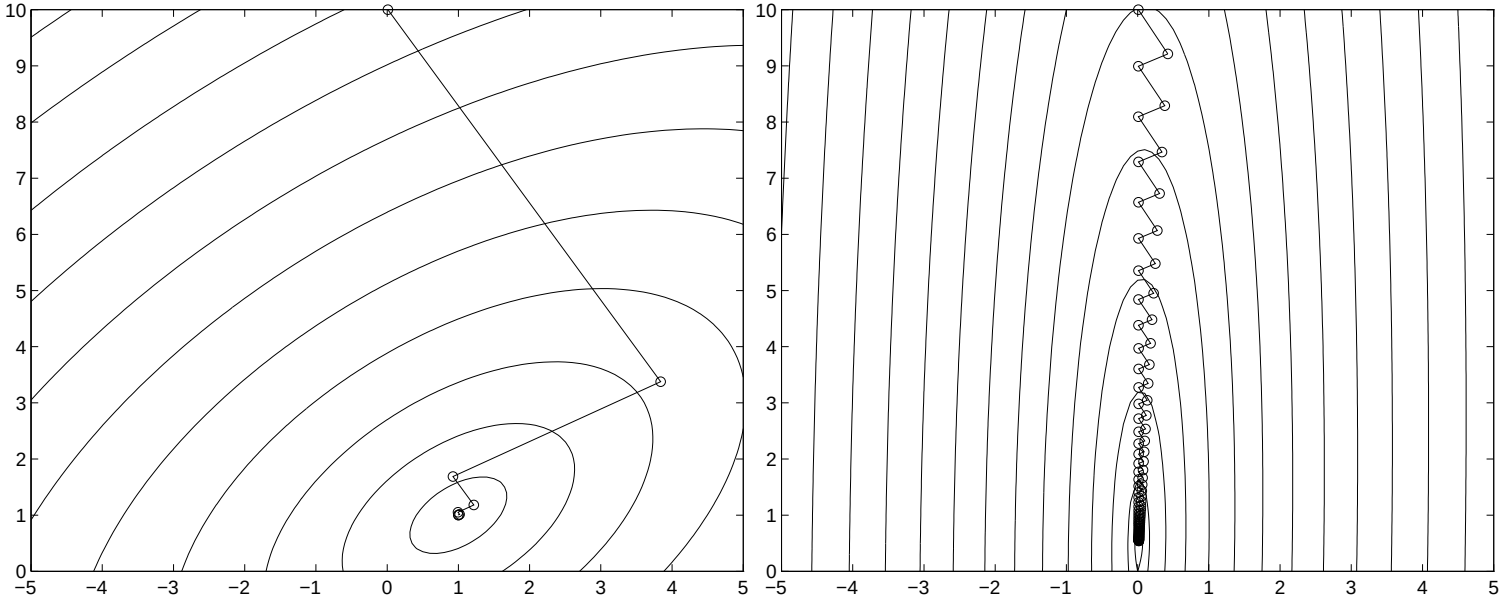
- Block relaxation: simultaneous relaxation of several equations by the same averaged amount.
- "Scheibenrelaxation":
  1. either choosing search directions related to eigenfunctions on subdomains.
  2. or solving directly small subproblems for low modes by relaxation.

  "Es ist zweckmässig, für einen gegebenen Operator eine Sammlung von Scheiben anzulegen."

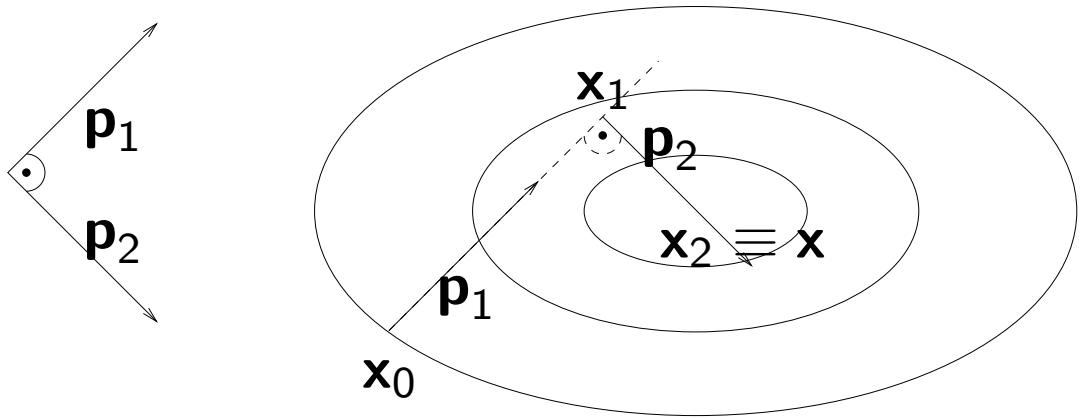
  These are precursors of **multigrid methods** and/or **domain decomposition**.
- Conjugate search directions: in that case, one can eliminate completely error components in the direction of each **p**, independent of the other directions.
- "Das n-Schritt Verfahren": the method of **conjugate gradients, (CG)**.

Problems of Steepest Descent

Examples of good and bad convergence of steepest descent



Idea: Try to use each search direction only once!



A Direct Method

Given N orthogonal search directions $\mathbf{p}_n$ how far do we have to go in

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \alpha_n \mathbf{p}_n$$

in order to never have to go again into this direction ?
Until the search direction is orthogonal to the direction where the solution lies,

$$\mathbf{p}_n^T \mathbf{e}_{n+1} = 0.$$

This determines the parameter α_n:

$$\begin{aligned} \mathbf{p}_n^T \mathbf{e}_{n+1} &= \mathbf{p}_n^T (\mathbf{x} - \mathbf{x}_{n+1}) \\ &= \mathbf{p}_n^T (\mathbf{x} - (\mathbf{x}_n + \alpha_n \mathbf{p}_n)) \\ &= \mathbf{p}_n^T \mathbf{e}_n - \alpha_n \mathbf{p}_n^T \mathbf{p}_n \;=\; 0 \end{aligned}$$

and thus

$$\alpha_n = \frac{\mathbf{p}_n^T \mathbf{e}_n}{\mathbf{p}_n^T \mathbf{p}_n}$$

This algorithm finds the solution in at most N steps.
BUT...

Conjugate Search Directions

We do not know $\mathbf{e}_n$! But we know $A\mathbf{e}_n$:

$$A\mathbf{e}_n = A\mathbf{x} - A\mathbf{x}_n = \mathbf{b} - A\mathbf{x}_n = \mathbf{r}_n$$

Hence if we have a set of A-orthogonal (conjugate) search directions $\mathbf{p}_n$, the algorithm

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \alpha_n \mathbf{p}_n$$

can use the A inner product to determine α_n:

$$\begin{aligned} 0 &= \mathbf{p}_n^T A\mathbf{e}_{n+1} = \mathbf{p}_n^T A(\mathbf{x} - \mathbf{x}_{n+1}) = \mathbf{p}_n^T A(\mathbf{x} - (\mathbf{x}_n + \alpha_n \mathbf{p}_n)) \\ &= \mathbf{p}_n^T A(\mathbf{x} - \mathbf{x}_n - \alpha_n \mathbf{p}_n) = \mathbf{p}_n^T A\mathbf{e}_n - \alpha_n \mathbf{p}_n^T A\mathbf{p}_n \\ &= \mathbf{p}_n^T \mathbf{r}_n - \alpha_n \mathbf{p}_n^T A\mathbf{p}_n \end{aligned}$$

Hence with

$$\alpha_n = \frac{\mathbf{p}_n^T \mathbf{r}_n}{\mathbf{p}_n A\mathbf{p}_n}$$

we obtain the solution in at most N computable steps!

Conjugate Gradient Algorithm

Suppose we have linearly independent vectors $\mathbf{r}_0, \ldots, \mathbf{r}_k$,

$$R_n := \text{span}\{\mathbf{r}_0, \ldots, \mathbf{r}_n\},$$

and an A orthogonal basis for R_n given by search directions $\mathbf{p}_0, \ldots, \mathbf{p}_n$,

$$P_n := \text{span}\{\mathbf{p}_0, \ldots, \mathbf{p}_n\} \quad (R_n \equiv P_n).$$

Using $\mathbf{p}_j$ in the algorithm of conjugate search directions, we have

$$\mathbf{p}_j^T A\mathbf{e}_{n+1} = 0 \quad \forall j \leq n.$$

But this implies

$$\mathbf{p}_j^T A\mathbf{e}_{n+1} = \mathbf{p}_j^T A(\mathbf{x} - \mathbf{x}_{n+1}) = \mathbf{p}_j^T(\mathbf{b} - A\mathbf{x}_{n+1}) = \mathbf{p}_j^T \mathbf{r}_{n+1} = 0$$

and hence the new residual $\mathbf{r}_{n+1}$ is orthogonal to all search directions used so far.

Conjugate Gradient Algorithm Continued

Since $\mathbf{p}_j^T \mathbf{r}_{n+1} = 0$, the new residual is also orthogonal to $P_n \equiv R_n$, and thus

$$\mathbf{r}_j^T \mathbf{r}_{n+1} = 0 \quad \forall j \leq n.$$

On the other hand,

$$\mathbf{r}_{n+1} = \mathbf{b} - A\mathbf{x}_{n+1} = \mathbf{b} - A\mathbf{x}_n - \alpha_n A\mathbf{p}_n = \mathbf{r}_n - \alpha_n A\mathbf{p}_n$$

which shows that $\mathbf{r}_{n+1}$ is a linear combination of $\mathbf{r}_n$ and $A\mathbf{p}_n$. Since $\mathbf{p}_n$ is in $P_n \equiv R_n$, the new residual is in $R_n \cup AR_n$ and thus

$$R_{n+1} = R_n \cup AR_n$$

By induction, we have

$$R_n (\equiv P_n) = \text{span}\{\mathbf{r}_0, A\mathbf{r}_0, A^2\mathbf{r}_0, \ldots, A^n\mathbf{r}_0\}$$

a **Krylov space**!

Conjugate Gradient Algorithm

Since $\mathbf{r}_{n+1}$ is orthogonal to $R_n \equiv P_n$, and $P_n = P_{n-1} \cup AP_{n-1}$, $\mathbf{r}_{n+1}$ is A-orthogonal to P_{n-1}. Therefore, to obtain a new search direction $\mathbf{p}_{n+1}$, it suffices to take $\mathbf{r}_{n+1}$ and make it A-orthogonal to $\mathbf{p}_n$, i.e.

$$\mathbf{p}_{n+1} := \mathbf{r}_{n+1} - \frac{\mathbf{r}_{n+1}^T A \mathbf{p}_n}{\mathbf{p}_n^T A \mathbf{p}_n} \mathbf{p}_n$$

Conjugate gradient algorithm:

```
guess x;
r=b-A*x; p=r;
for n=1:N,
  al=p'*r/(p'*A*p);        % remove error in direction p
  x=x+al*p;
  r=b-A*x;
  p=r-r'*A*p/(p'*A*p)*p;   % determine new direction p
end;
```

Minimization Property of Conjugate Gradients

CG finds in $R_n = \text{span}\{\mathbf{r}_0, A\mathbf{r}_0, \ldots, A^n\mathbf{r}_0\} + \mathbf{x}_0$ the best approximation to the solution in the A-norm, i.e.

$$\mathbf{x}_{n+1} = \sum_{j=0}^{n} \beta_j A^j \mathbf{r}_0 + \mathbf{x}_0$$

such that $\mathbf{e}_{n+1}$ is A orthogonal to R_n, $\mathbf{e}_{n+1}^T A\mathbf{y} = 0$, $\forall \mathbf{y} \in R_n$. But this implies that

$$\mathbf{x}_{n+1} = \tilde{p}_n(A)(\mathbf{b} - A\mathbf{x}_0) + \mathbf{x}_0 = \tilde{p}_n(A)A(\mathbf{x} - \mathbf{x}_0) + \mathbf{x}_0$$

and by subtracting both sides from $\mathbf{x}$, we find

$$\begin{aligned} \mathbf{x} - \mathbf{x}_{n+1} &= -\tilde{p}_n(A)A(\mathbf{x} - \mathbf{x}_0) + \mathbf{x} - \mathbf{x}_0 \\ &= (-\tilde{p}_n(A)A + I)(\mathbf{x} - \mathbf{x}_0) = p_{n+1}(A)(\mathbf{x} - \mathbf{x}_0) \end{aligned}$$

and hence $\mathbf{e}_{n+1} = p(A)_{n+1}\mathbf{e}_0$, where the residual polynomial satisfies $p_{n+1}(0) = 1$. Hence CG finds the residual polynomial $p_n(A)$ such that the error $\mathbf{e}_n$ is minimized in the A-norm.

Lanczos 1950, 1952

"This seems to be a very elementary problem without deeper meaning. However, one meets this task again and again in the electric industry and in all kinds of oscillation problems. A short while ago, I found a rather elegant solution. The reason why I am strongly drawn to such approximation mathematics problems is not the practical applicability of the solution, but rather the fact that a very economical solution is possible only when it is very adequate. To obtain a solution in very few steps means nearly always that one has found a way that does justice to the inner nature of the problem." (**Cornelius Lanczos** in a letter to Albert Einstein on March 9, 1947)

"Your remark on the importance of adapted approximation methods makes very good sense to me, and I am convinced that this is a fruitful mathematical aspect, and not just a utilitarian one." (**Einsteins** reply to Lanczos on March 18, 1947)

MIRES and SYMMLQ 1975

SIAM J. NUMER. ANAL.
Vol. 12, No. 4, September 1975

SOLUTION OF SPARSE INDEFINITE SYSTEMS OF LINEAR EQUATIONS*

C. C. PAIGE† AND M. A. SAUNDERS‡

Abstract. The method of conjugate gradients for solving systems of linear equations with a symmetric positive definite matrix A is given as a logical development of the Lanczos algorithm for tridiagonalizing A. This approach suggests numerical algorithms for solving such systems when A is symmetric but indefinite. These methods have advantages when A is large and sparse.

Two Krylov methods for symmetric indefinite problems:

- SYMMLQ: algorithm like CG, based on the Lanczos process and the LQ factorization
- MINRES: minimizes the residual

Both based on short term recurrences

FOM and GMRES 1986

SIAM J. SCI. STAT. COMPUT.
Vol. 7, No. 3, July 1986

011

GMRES: A GENERALIZED MINIMAL RESIDUAL ALGORITHM FOR SOLVING NONSYMMETRIC LINEAR SYSTEMS*

YOUCEF SAAD† AND MARTIN H. SCHULTZ†

Abstract. We present an iterative method for solving linear systems, which has the property of minimizing at every step the norm of the residual vector over a Krylov subspace. The algorithm is derived from the Arnoldi process for constructing an l_2-orthogonal basis of Krylov subspaces. It can be considered as a generalization of Paige and Saunders' MINRES algorithm and is theoretically equivalent to the Generalized Conjugate Residual (GCR) method and to ORTHODIR. The new algorithm presents several advantages over GCR and ORTHODIR.

Two Krylov methods for general matrices:

- FOM: algorithm computing approximations by making the new residual orthogonal to the current Krylov space
- GMRES: generalized minimum residual method, based on Arnoldi, mathematically equivalent to GCR (PhD of Elman 1982), ORTHOMIN (Vinsome 1976), ORTHODIR (Jea and Young 1980)

Both based on long term recurrences

Numer. Math. 60, 315–339 (1991)

Numerische Mathematik

QMR: a quasi-minimal residual method for non-Hermitian linear systems*

Roland W. Freund[1, 2] **and Noël M. Nachtigal**[3]

[1] RIACS, Mail Stop Ellis Street, NASA Ames Research Center, Moffett Field, CA 94035, USA

[2] Institut für Angewandte Mathematik und Statistik, Universität Würzburg, W-8700 Würzburg, Federal Republic of Germany

[3] Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received February 19, 1991

Summary. The biconjugate gradient (BCG) method is the "natural" generalization of the classical conjugate gradient algorithm for Hermitian positive definite matrices to general non-Hermitian linear systems. Unfortunately, the original BCG algorithm is susceptible to possible breakdowns and numerical instabilities. In this paper, we present a novel BCG-like approach, the quasi-minimal residual (QMR) method, which overcomes the problems of BCG. An implementation of QMR based on a look-ahead version of the nonsymmetric Lanczos algorithm is proposed. It is shown how BCG iterates can be recovered stably from the QMR process. Some further properties of the QMR approach are given and an error bound is presented. Finally, numerical experiments are reported.

Bi-CGSTAB 1992

SIAM J. SCI. STAT. COMPUT.
Vol. 13, No. 2, pp. 631–644, March 1992

BI-CGSTAB: A FAST AND SMOOTHLY CONVERGING VARIANT OF BI-CG FOR THE SOLUTION OF NONSYMMETRIC LINEAR SYSTEMS*

H. A. VAN DER VORST†

Abstract. Recently the Conjugate Gradients-Squared (CG-S) method has been proposed as an attractive variant of the Bi-Conjugate Gradients (Bi-CG) method. However, it has been observed that CG-S may lead to a rather irregular convergence behaviour, so that in some cases rounding errors can even result in severe cancellation effects in the solution. In this paper, another variant of Bi-CG is proposed which does not seem to suffer from these negative effects. Numerical experiments indicate also that the new variant, named Bi-CGSTAB, is often much more efficient than CG-S.

- Based on earlier work on Bi-CG by Fletcher 1976, and CG-S and IDR by Sonneveld 1989
- Contructs two sequences in Krylov spaces from A and A^T

Uses only short term recurrences

The Name Krylov Space ?

ИЗВЕСТИЯ АКАДЕМИИ НАУК СССР. 1931

BULLETIN DE L'ACADÉMIE DES SCIENCES DE L'URSS

Classe des sciences mathématiques et naturelles

Отделение математических и естественных наук

Метода Лапласа видимо не удовлетворяла Леверрье, так как в своей статье: «Sur les variations séculaires des éléments des orbites», помещенной в прибавлении к «Connaissance des Temps» за 1843 г., следовательно напечатанной в 1839 или 1840 г., Леверрье излагает свою методу составления векового уравнения. Эта метода включена затем и в его «Recherches Astronomiques» (гл. IX), помещенных в «Annales de l'Observatoire Impérial de Paris» t. II, 1856 г.

Главное неудобство вида (14) векового уравнения для *численного* его решения происходит вследствие того, что члены вида $a_{ii} - \lambda^2$ стоят в диагонали определителя. Это заставило знаменитого математика Якоби поместить в т. XXX журнала «Crelle» составленную им в 1845 г. обширную статью под заглавием: «Ueber ein leichtes Verfahren, die in der Theorie der Secularstörungen vorkommenden Gleichungen numerisch aufzulösen». Эта статья вошла в т. VII собрания сочинений Якоби.

Topic is Second Order ODEs

Krylov explains a method to solve the system of ODEs

$$\left.\begin{aligned}
\ddot{q}_1 &= a_{11}q_1 + a_{12}q_2 + a_{13}q_3 + \cdots + a_{1k}q_k \\
\ddot{q}_2 &= a_{21}q_1 + a_{22}q_2 + a_{23}q_3 + \cdots + a_{2k}q_k \\
\ddot{q}_3 &= a_{31}q_1 + a_{32}q_2 + a_{33}q_3 + \cdots + a_{3k}q_k \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\ddot{q}_k &= a_{k1}q_1 + a_{k2}q_2 + a_{k3}q_3 + \cdots + a_{kk}q_k
\end{aligned}\right\} \quad \ldots\ldots\ldots(11)$$

Для такой системы при обычном теперь изложении ищут частное решение вида:

$$q_1 = C_1 e^{\lambda t}; \quad q_2 = C_2 e^{\lambda t}; \quad q_3 = C_3 e^{\lambda t} \ldots q_k = C_k e^{\lambda t} \ldots\ldots(12)$$

In order to find the analytical solution, one needs to obtain the eigenvalues λ of the matrix of the system.

Computing Eigenvalues

Для совместности этих уравнений необходимо, чтобы было

$$\Delta(\lambda) = \begin{vmatrix} (a_{11} - \lambda^2) & a_{12} & a_{13} & \dots & a_{1k} \\ a_{21} & (a_{22} - \lambda^2) & a_{23} & \dots & a_{2k} \\ a_{31} & a_{32} & (a_{33} - \lambda^2) & \dots & a_{3k} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ a_{k1} & a_{k2} & a_{k3} & \dots & (a_{kk} - \lambda^2) \end{vmatrix} = 0 \dots (14)$$

$$\alpha^k - (a_{11} + a_{22} + a_{33} + \dots + a_{kk})\,\alpha^{k-1}. \quad \dots\dots\dots (21)$$

Отсюда видно, что сумма корней уравнения (19) будет:

$$s_1 = \alpha_1 + \alpha_2 + \dots + \alpha_k = a_{11} + a_{22} + a_{33} + \dots + a_{kk} \quad \dots (22)$$

Powers of Matrices

Krylov now explains the method of Leverrier

$$\left.\begin{aligned} q_1^{IV} &= a_{11}\ddot{q}_1 + a_{12}\ddot{q}_2 + \cdots + a_{1k}\ddot{q}_k \\ q_2^{IV} &= a_{12}\ddot{q}_1 + a_{22}\ddot{q}_2 + \cdots + a_{2k}\ddot{q}_k \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ q_k^{IV} &= a_{k1}\ddot{q}_1 + a_{k2}\ddot{q}_2 + \cdots + a_{2k}\ddot{q}_k \end{aligned}\right\} \quad \ldots\ldots\ldots(23)$$

$$\left.\begin{aligned} q_1^{IV} &= b_{11}q_1 + b_{12}q_2 + b_{13}q_3 + \cdots + b_{1k}q_k \\ q_2^{IV} &= b_{21}q_1 + b_{22}q_2 + b_{23}q_3 + \cdots + b_{2k}q_k \\ q_3^{IV} &= b_{31}q_1 + b_{32}q_2 + b_{33}q_3 + \cdots + b_{3k}q_k \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ q_k^{IV} &= b_{k1}q_1 + b_{k2}q_2 + b_{k3}q_3 + \cdots + b_{kk}q_k \end{aligned}\right\} \quad \ldots\ldots\ldots(26)$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1k} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2k} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{k1} & a_{k2} & a_{k3} & \cdots & a_{kk} \end{vmatrix} \cdot \begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1k} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2k} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{k1} & a_{k2} & a_{k3} & \cdots & a_{kk} \end{vmatrix} = \begin{vmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1k} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{2k} \\ b_{31} & b_{32} & b_{33} & \cdots & b_{3k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{k1} & b_{k2} & b_{k3} & \cdots & b_{kk} \end{vmatrix} \quad (27)$$

Powers of Matrices

$$s_2 = \alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \cdots + \alpha_k^2 = b_{11} + b_{22} + b_{33} + \cdots + b_{kk} \quad \ldots(30)$$

$$\begin{vmatrix} b_{11} & b_{12} & \ldots & b_{1k} \\ b_{21} & b_{22} & \ldots & b_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ b_{k1} & b_{k2} & \ldots & b_{kk} \end{vmatrix} \cdot \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1k} \\ a_{21} & a_{22} & \ldots & a_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ a_{k1} & a_{k2} & \ldots & a_{kk} \end{vmatrix} = \begin{vmatrix} c_{11} & c_{12} & \ldots & c_{1k} \\ c_{21} & c_{22} & \ldots & c_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ c_{k1} & c_{k2} & \ldots & c_{kk} \end{vmatrix} \quad \ldots\ldots(32)$$

$$s_3 = \alpha_1^3 + \alpha_2^3 + \alpha_3^3 + \cdots + \alpha_k^3 = c_{11} + c_{22} + c_{33} + \cdots + c_{kk}. \quad \ldots(33)$$

Продолжая поступать таким образом получим последовательно значения

$$\left.\begin{aligned} s_4 &= \alpha_1^4 + \alpha_2^4 + \alpha_3^4 + \cdots + \alpha_k^4 \\ s_5 &= \alpha_1^5 + \alpha_2^5 + \alpha_3^5 + \cdots + \alpha_k^5 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ s_k &= \alpha_1^k + \alpha_2^k + \alpha_3^k + \cdots + \alpha_k^k \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots(34)$$

Using Newton's Formulas

Krylov arrives at the general characteristic polynomial

$$\Delta(\alpha) = \alpha^k + B_1 \alpha^{k-1} + B_2 \alpha^{k-2} + \cdots + B_k = 0. \quad \ldots\ldots (35)$$

using the s_i and the well known Newton formulas

Зная значения

$$s_1,\ s_2,\ s_3 \ldots s_k$$

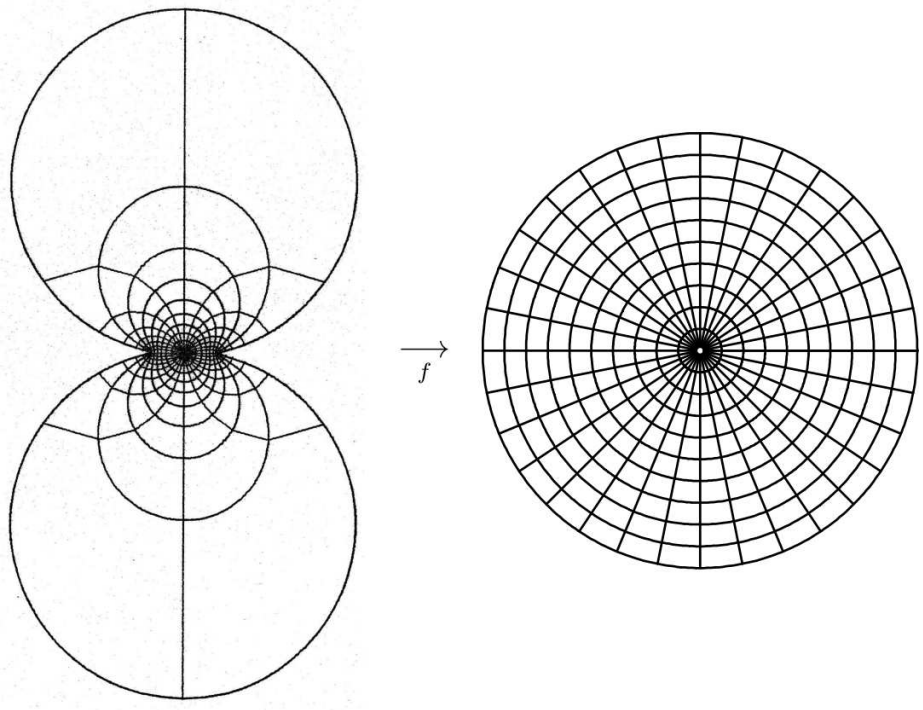
найдем коэфициенты $B_1, B_2 \ldots B_k$ по известным формулам Ньютона:

$$\left.\begin{aligned}
0 &= s_1 + B_1 \\
0 &= s_2 + B_1 s_1 + 2B_2 \\
0 &= s_3 + B_1 s_2 + B_2 s_1 + 3B_3 \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
0 &= s_{k-1} + B_1 s_{k-2} + B_2 s_{k-3} + \cdots + B_{k-2} s_1 + (k-1) B_{k-1} \\
0 &= s_k + B_1 s_{k-1} + B_2 s_{k-2} + \cdots + B_{k-1} s_1 + k B_k
\end{aligned}\right\} \ldots (36)$$

The origin of Krylov methods is not immediate!

Origin of Schwarz Methods

Riemann Mapping Theorem: "Zwei gegebene einfach zusammenhängende Flächen können stets so aufeinander bezogen werden, dass jedem Punkte der einen ein mit ihm stetig fortrückender Punkt entspricht...;"



(drawing M. Gutknecht 18.12.1975)

Proof: Riemann uses existence of harmonic functions (Dirichlet principle)

International Challenge

Find harmonic functions $\Delta u = 0$ on any domain Ω with prescribed boundary conditions $u = g$ for $(x, y) \in \partial\Omega$.

- Solution easy for circular domain (Poisson 1815):

$$u(r,\phi) = \frac{1}{2\pi} \int_0^{2\pi} \frac{1 - r^2}{1 - 2r\cos(\phi - \psi) + r^2} f(\psi)\, d\psi$$

- Solution also easy for rectangular domains (Fourier 1807): Fourier series

But existence of solutions of Laplace equation on arbitrary domains appears hopeless !

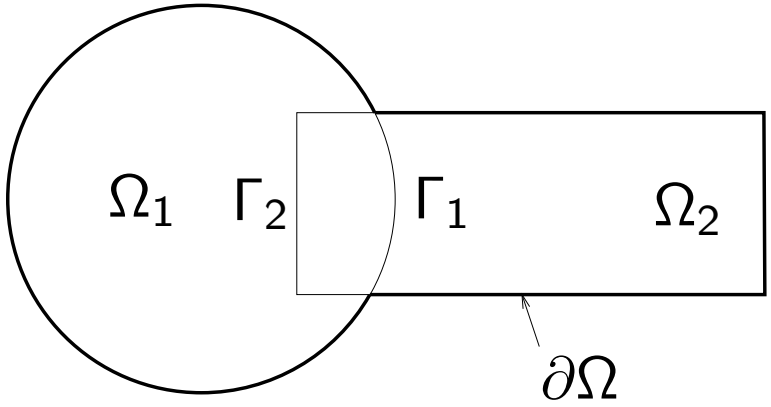
Proof without Dirichlet Principle

H.A. Schwarz (1870, Crelle 74, 1872) Über einen Grenzübergang durch alternierendes Verfahren

"Die unter dem Namen Dirichletsches Princip bekannte Schlussweise, welche in gewissem Sinne als das Fundament des von Riemann entwickelten Zweiges der Theorie der analytischen Functionen angesehen werden muss, unterliegt, wie jetzt wohl allgemein zugestanden wird, hinsichtlich der Strenge sehr begründeten Einwendungen, deren vollständige Entfernung meines Wissens den Anstrengungen der Mathematiker bisher nicht gelungen ist".

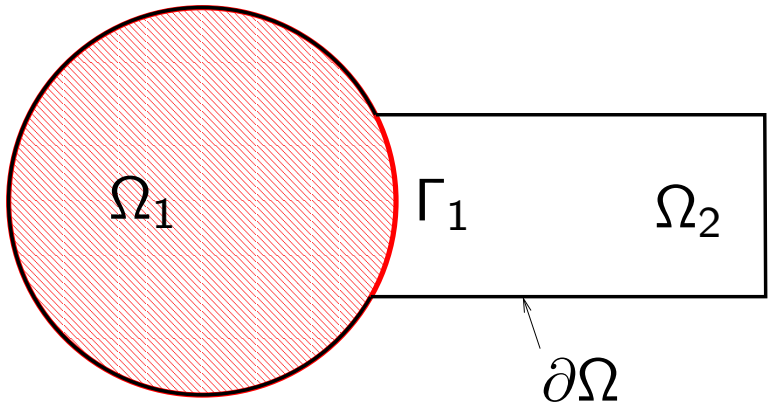
Classical Alternating Schwarz Method

Schwarz invents a method to proof that the infimum is attained: for a general domain $\Omega := \Omega_1 \cup \Omega_2$:



Classical Alternating Schwarz Method

Schwarz invents a method to proof that the infimum is attained: for a general domain $\Omega := \Omega_1 \cup \Omega_2$:

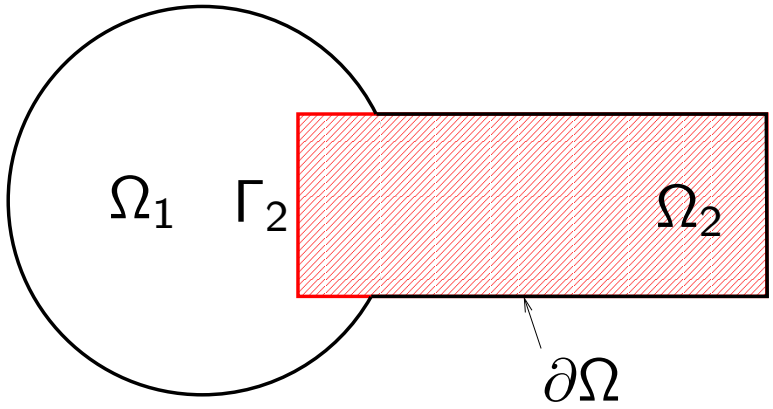


$$\begin{aligned} \Delta u_1^1 &= 0 && \text{in } \Omega_1 \\ u_1^1 &= g && \text{on } \partial\Omega \cap \overline{\Omega}_1 \\ u_1^1 &= u_2^0 && \text{on } \Gamma_1 \end{aligned}$$

solve on the disk $u_2^0 = 0$

Classical Alternating Schwarz Method

Schwarz invents a method to proof that the infimum is attained: for a general domain $\Omega := \Omega_1 \cup \Omega_2$:

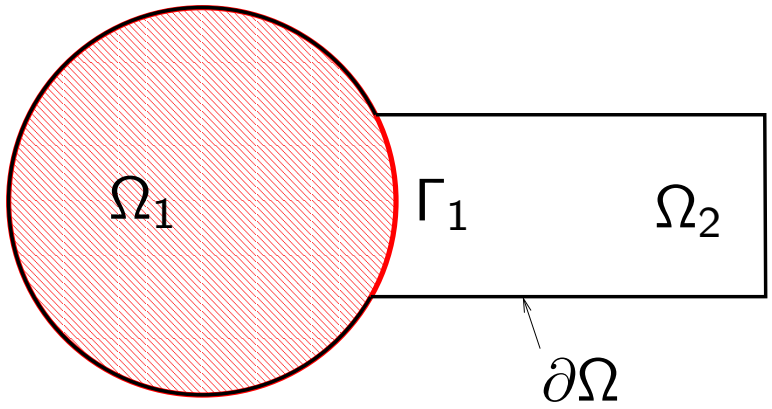


$$\begin{aligned} \Delta u_2^1 &= 0 && \text{in } \Omega_2 \\ u_2^1 &= g && \text{on } \partial\Omega \cap \overline{\Omega}_2 \\ u_2^1 &= u_1^1 && \text{on } \Gamma_2 \end{aligned}$$

solve on the rectangle

Classical Alternating Schwarz Method

Schwarz invents a method to proof that the infimum is attained: for a general domain $\Omega := \Omega_1 \cup \Omega_2$:

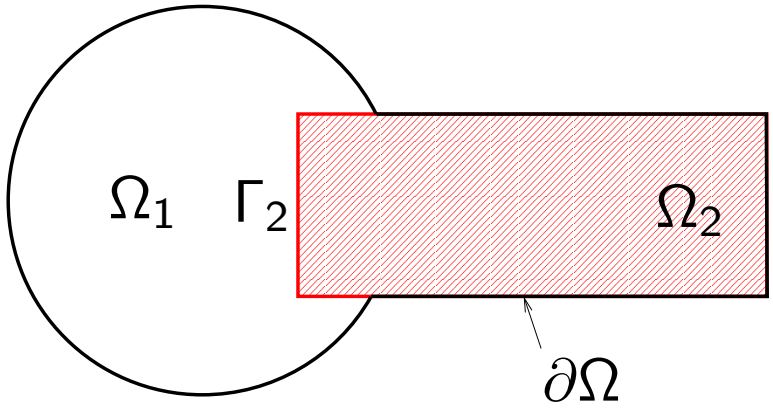


$$\begin{aligned} \Delta u_1^2 &= 0 && \text{in } \Omega_1 \\ u_1^2 &= g && \text{on } \partial\Omega \cap \overline{\Omega}_1 \\ u_1^2 &= u_2^1 && \text{on } \Gamma_1 \end{aligned}$$

solve on the disk

Classical Alternating Schwarz Method

Schwarz invents a method to proof that the infimum is attained: for a general domain $\Omega := \Omega_1 \cup \Omega_2$:

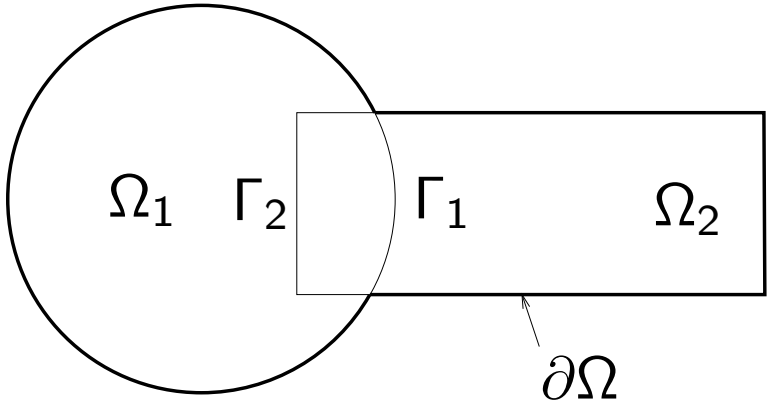


$$\begin{aligned} \Delta u_2^2 &= 0 && \text{in } \Omega_2 \\ u_2^2 &= g && \text{on } \partial\Omega \cap \overline{\Omega}_2 \\ u_2^2 &= u_1^2 && \text{on } \Gamma_2 \end{aligned}$$

solve on the rectangle

Classical Alternating Schwarz Method

Schwarz invents a method to proof that the infimum is attained: for a general domain $\Omega := \Omega_1 \cup \Omega_2$:



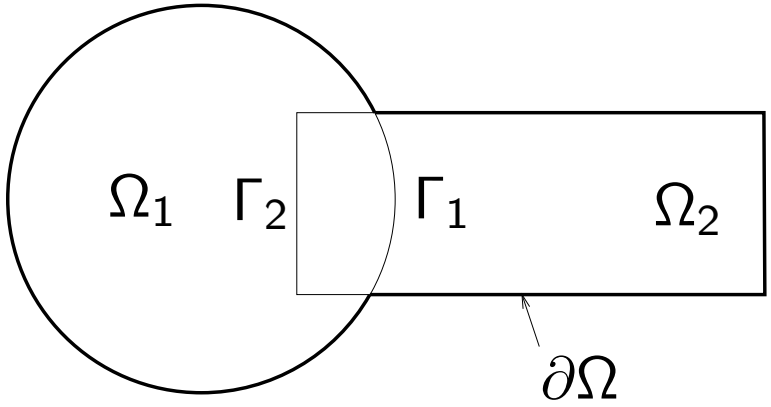
$$\begin{aligned} \Delta u_1^n &= 0 && \text{in } \Omega_1 \\ u_1^n &= g && \text{on } \partial\Omega \cap \overline{\Omega}_1 \\ u_1^n &= u_2^{n-1} && \text{on } \Gamma_1 \end{aligned} \qquad \begin{aligned} \Delta u_2^n &= 0 && \text{in } \Omega_2 \\ u_2^n &= g && \text{on } \partial\Omega \cap \overline{\Omega}_2 \\ u_2^n &= u_1^n && \text{on } \Gamma_2 \end{aligned}$$

solve on the disk

solve on the rectangle

Classical Alternating Schwarz Method

Schwarz invents a method to proof that the infimum is attained: for a general domain $\Omega := \Omega_1 \cup \Omega_2$:

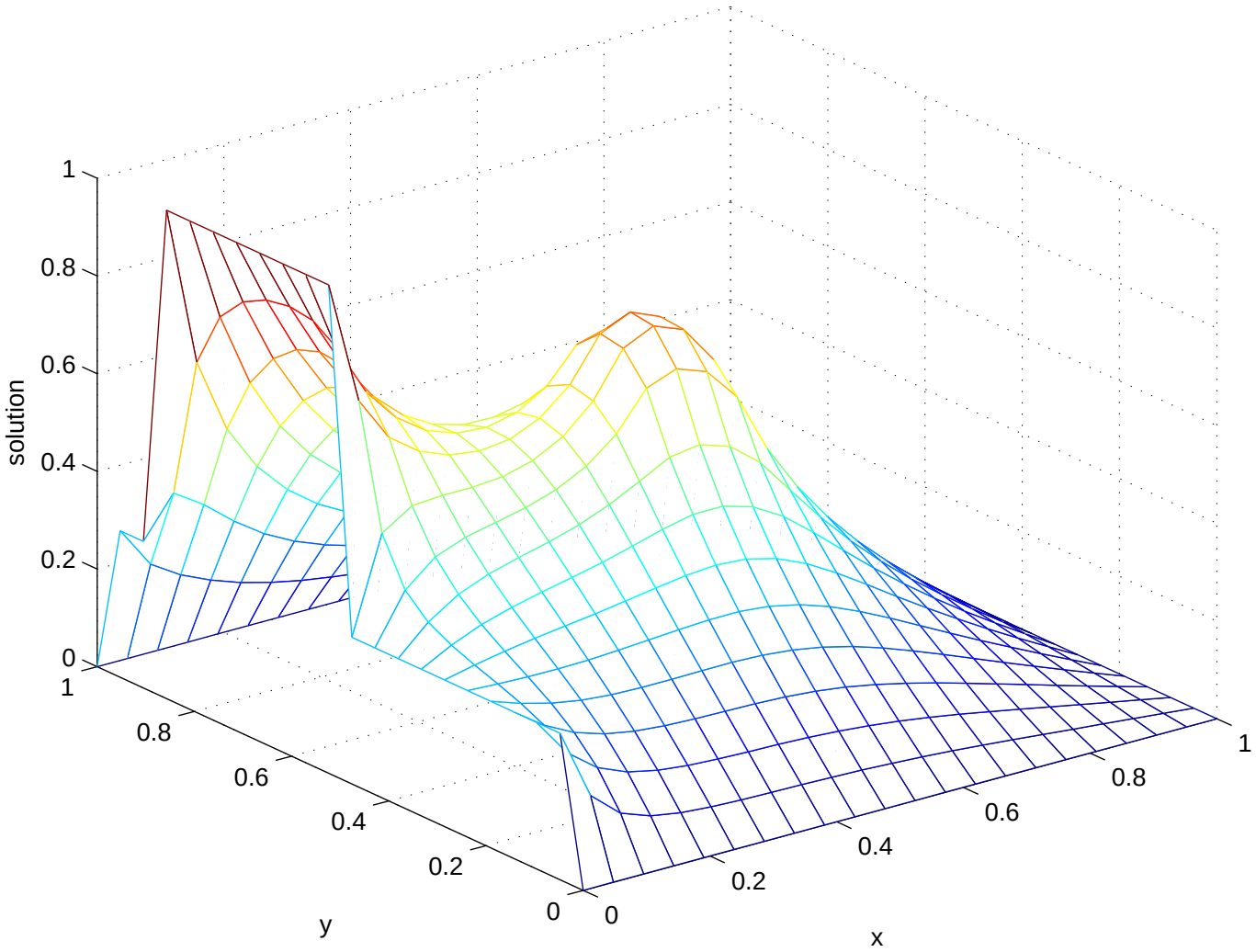


$$\begin{aligned} \Delta u_1^n &= 0 && \text{in } \Omega_1 \\ u_1^n &= g && \text{on } \partial\Omega \cap \overline{\Omega}_1 \\ u_1^n &= u_2^{n-1} && \text{on } \Gamma_1 \end{aligned} \qquad \begin{aligned} \Delta u_2^n &= 0 && \text{in } \Omega_2 \\ u_2^n &= g && \text{on } \partial\Omega \cap \overline{\Omega}_2 \\ u_2^n &= u_1^n && \text{on } \Gamma_2 \end{aligned}$$

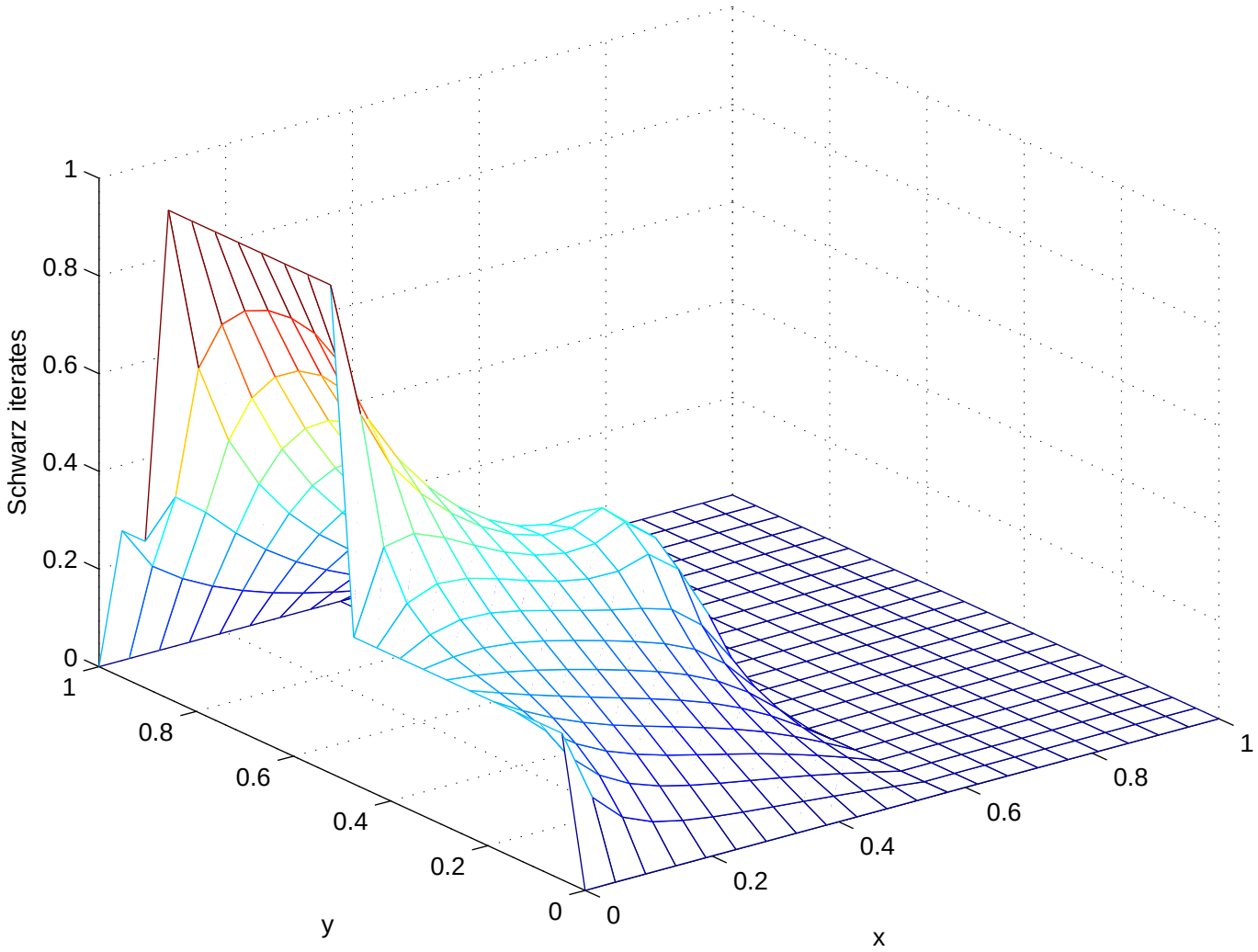
solve on the disk solve on the rectangle

- Schwarz proved convergence in 1869 using the maximum principle.

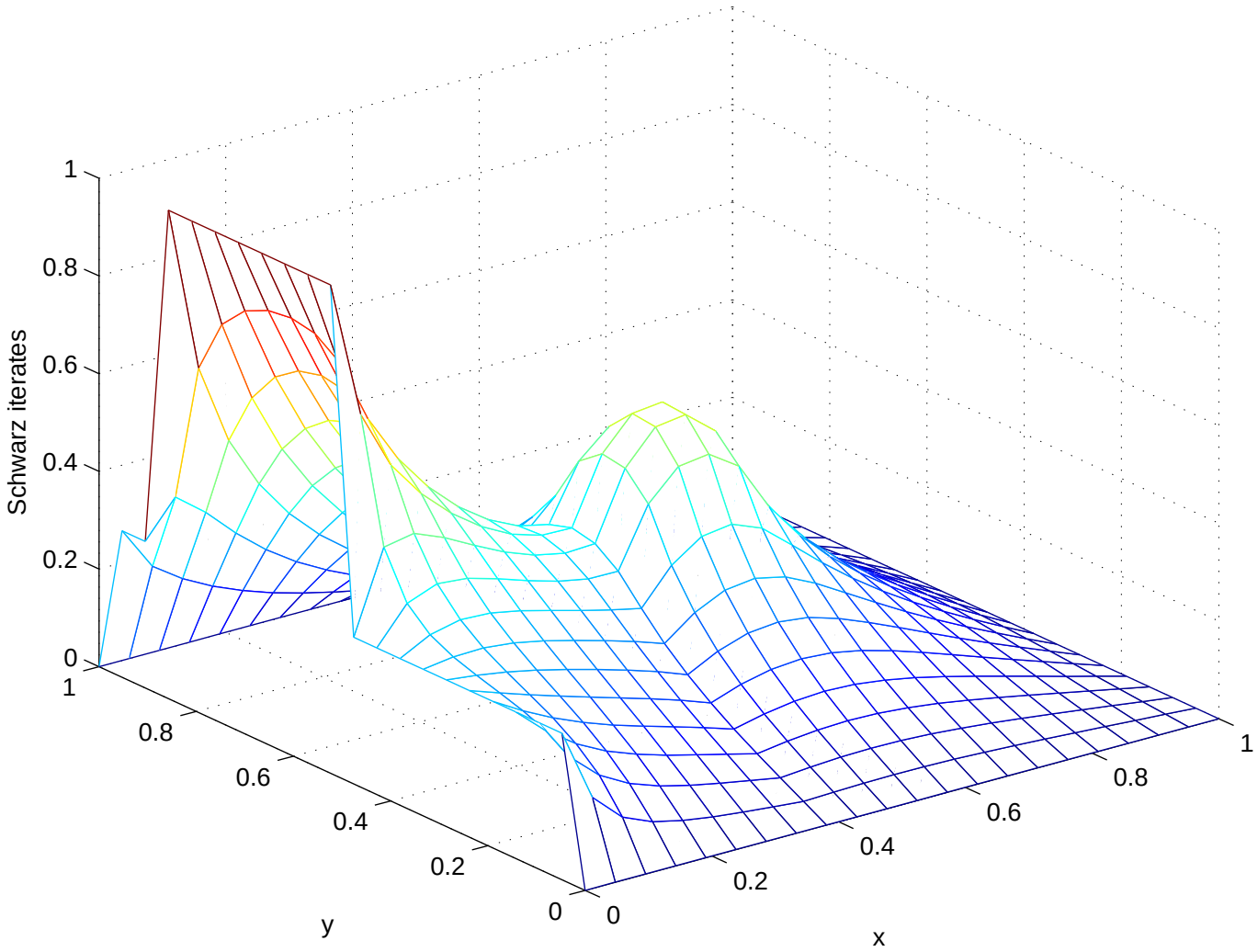
Example: Heating a Room



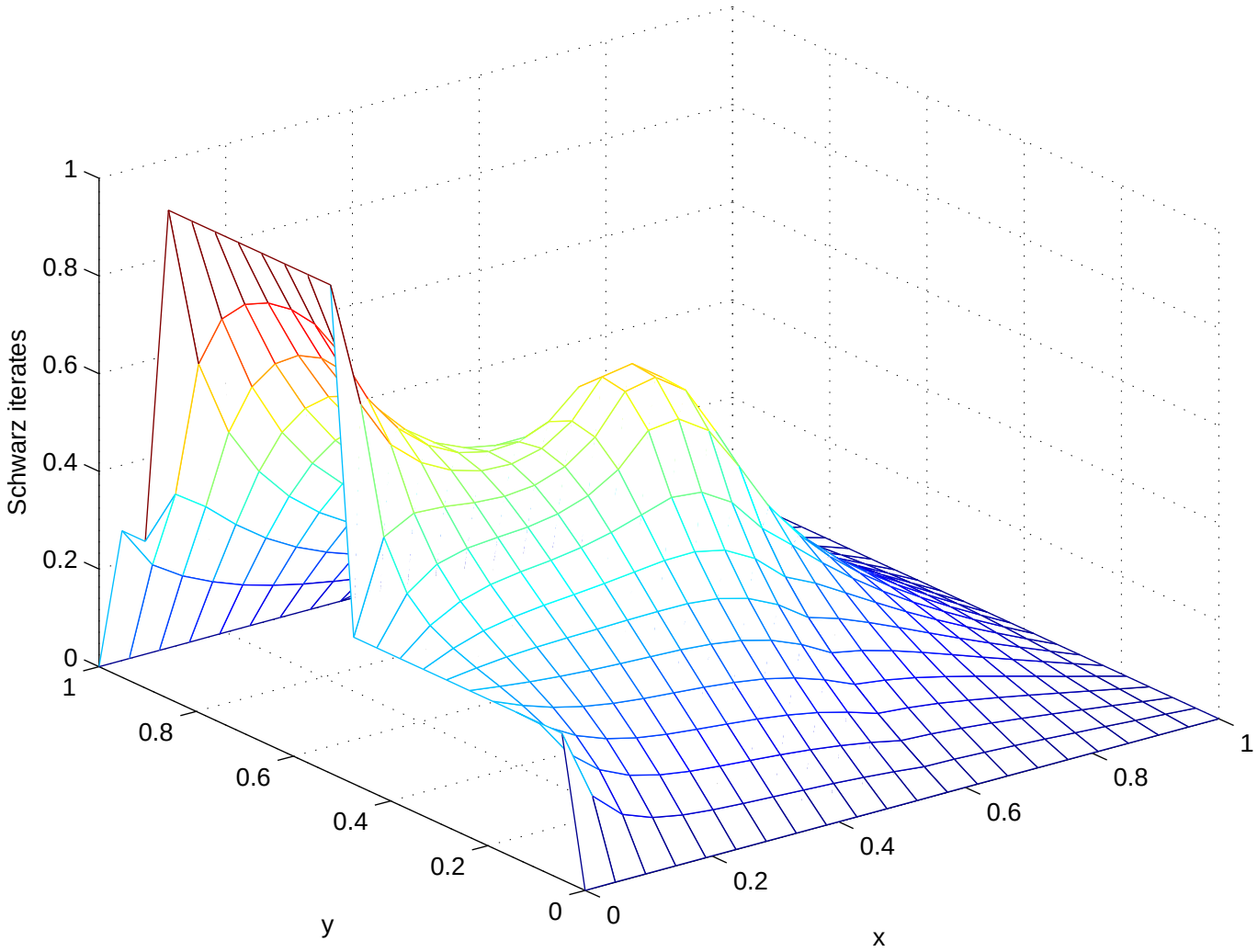
Iteration 1



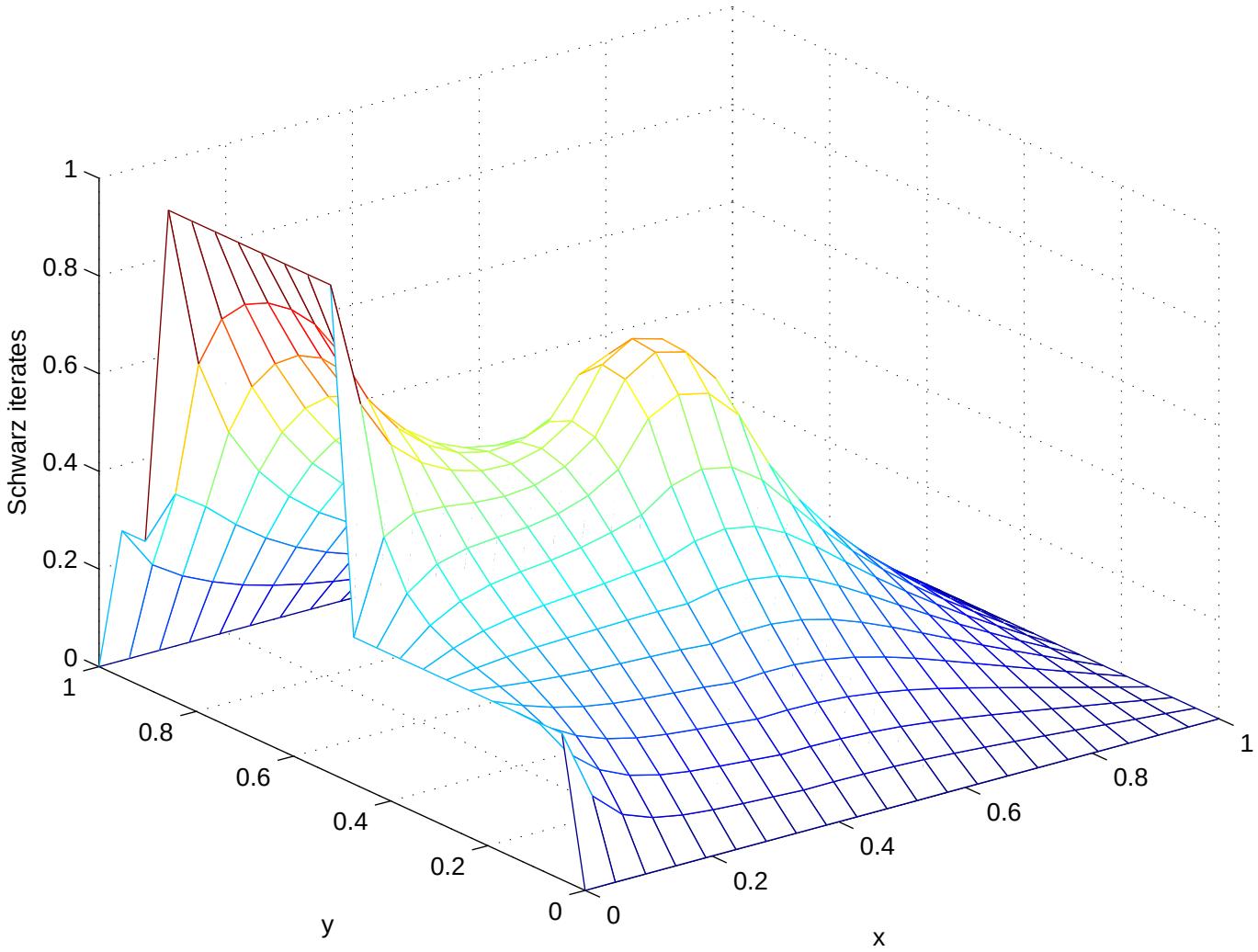
Iteration 2



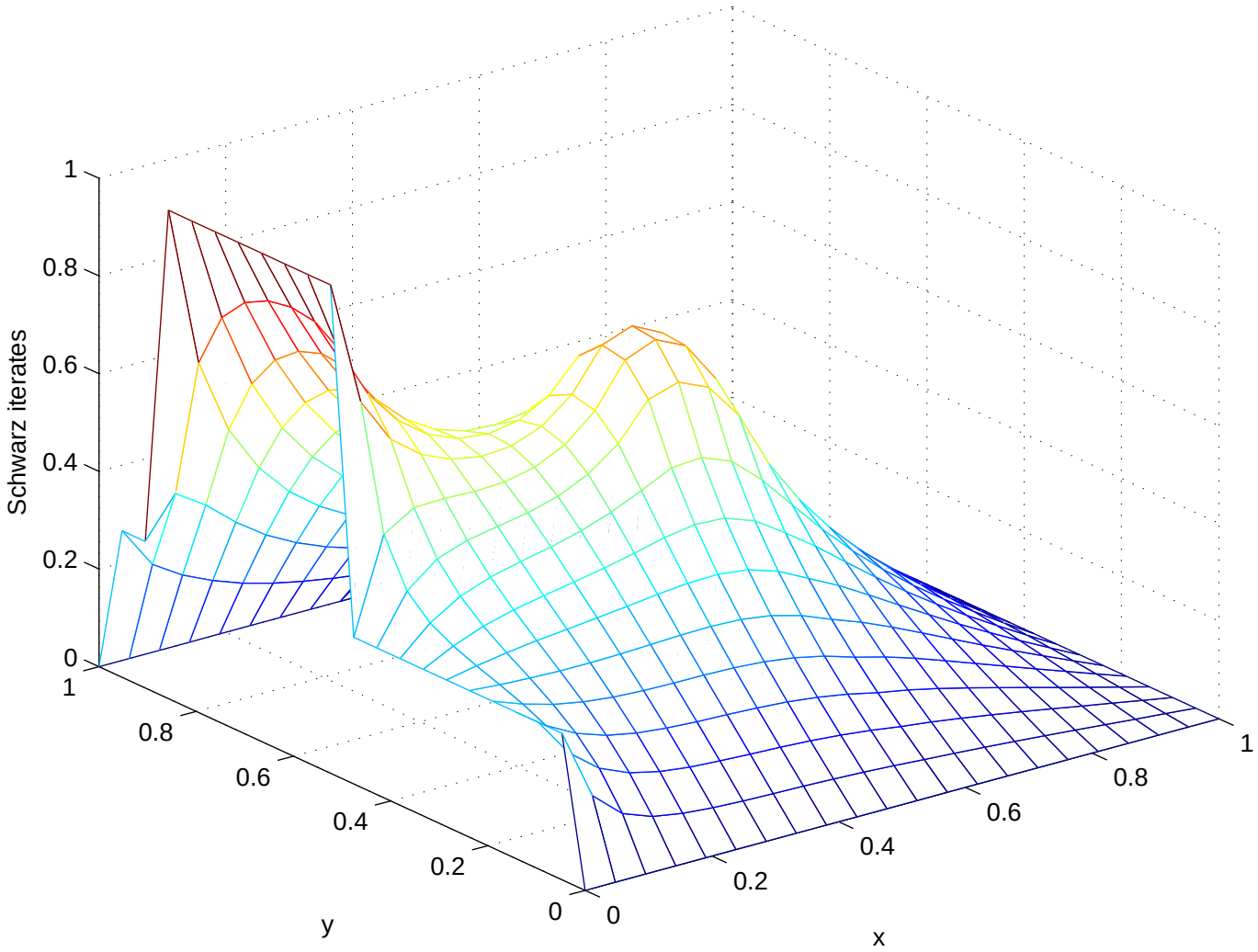
Iteration 3



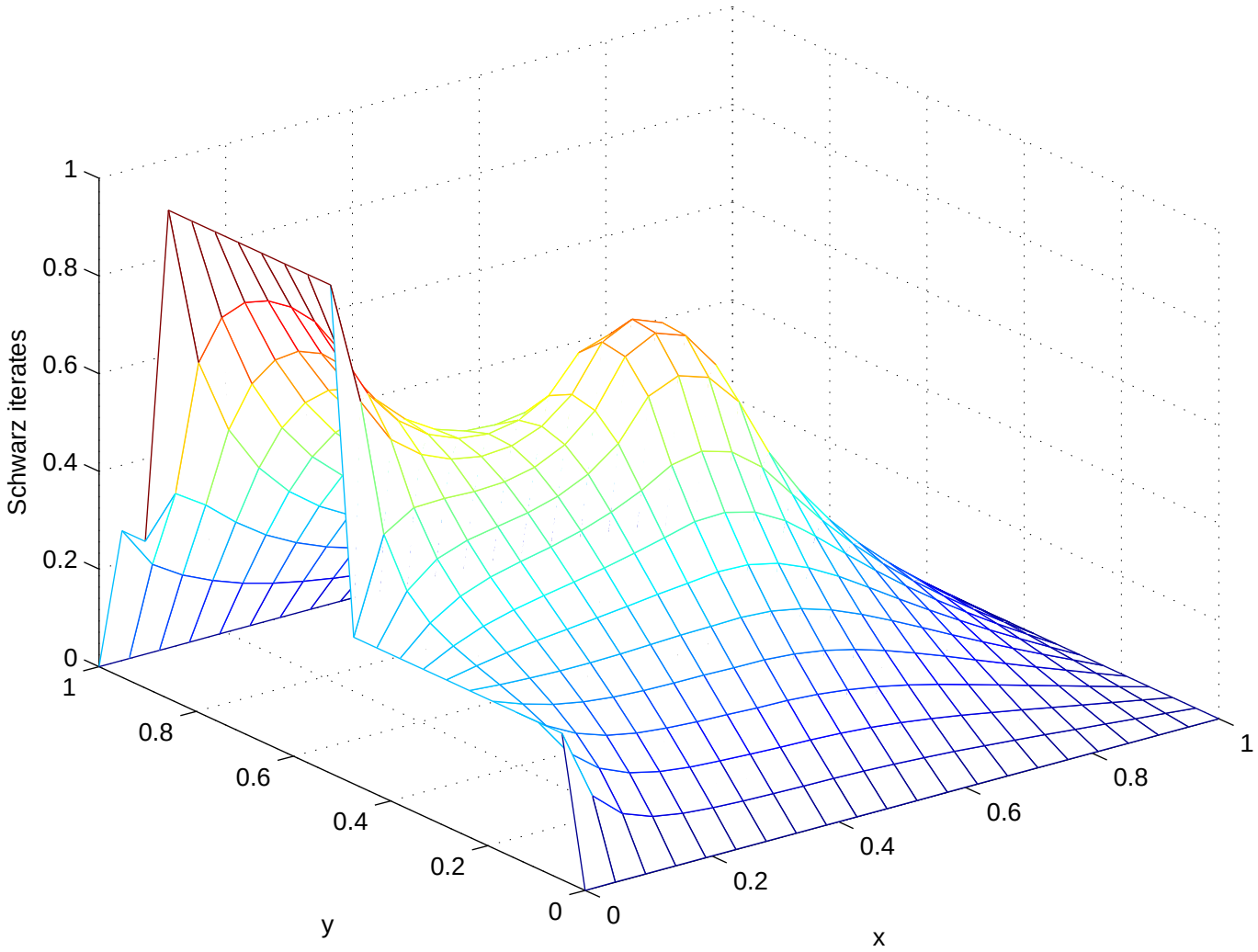
Iteration 4



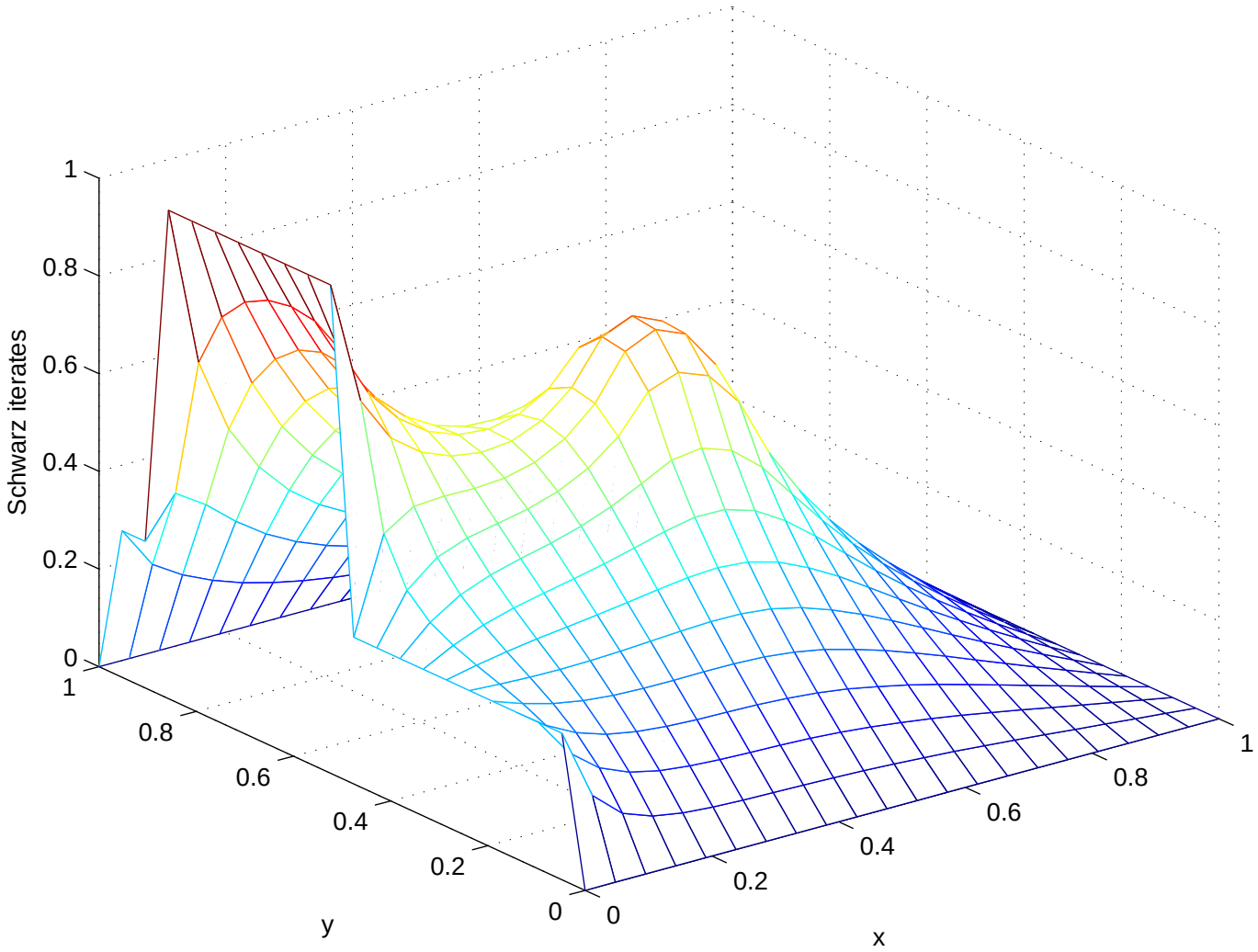
Iteration 5



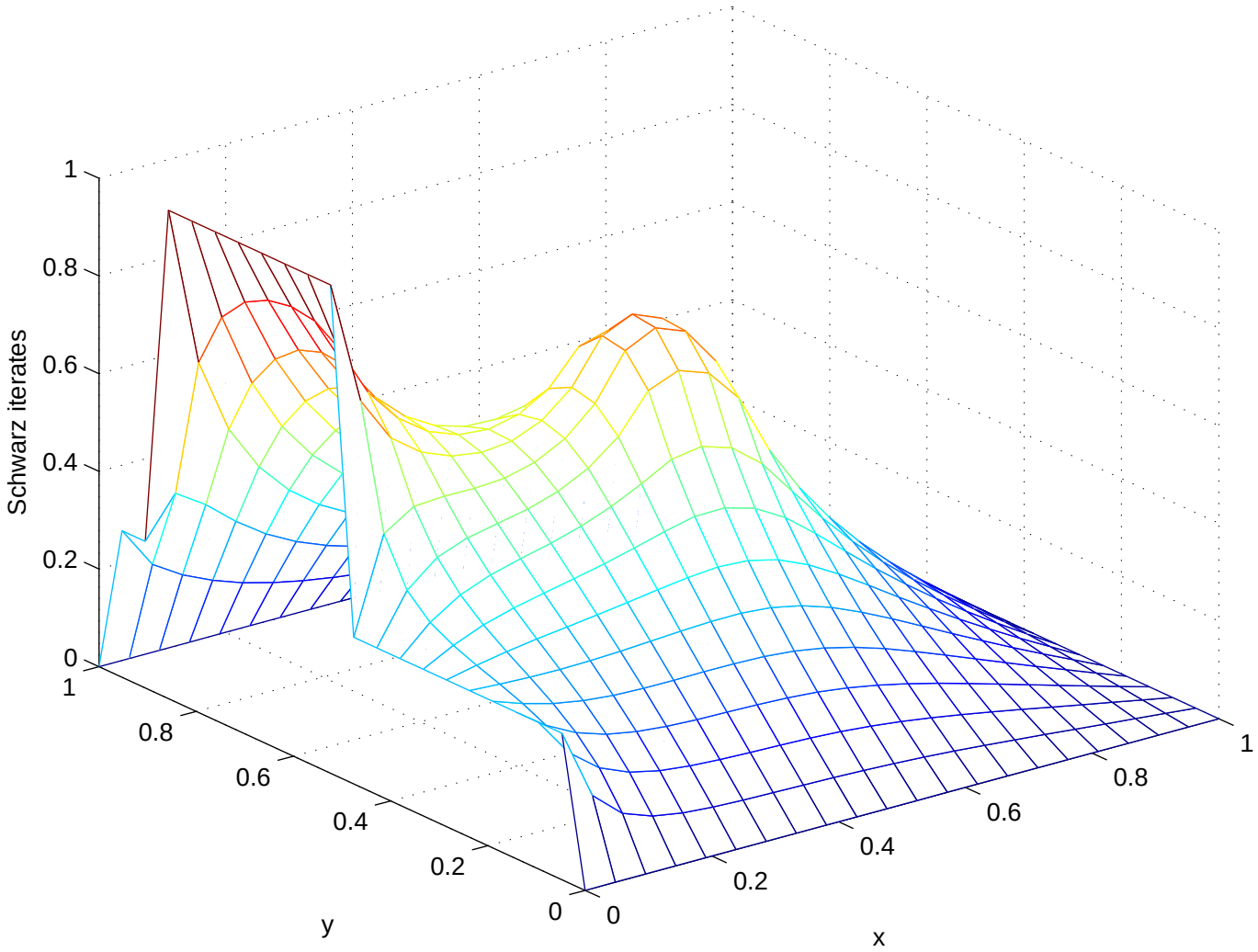
Iteration 6



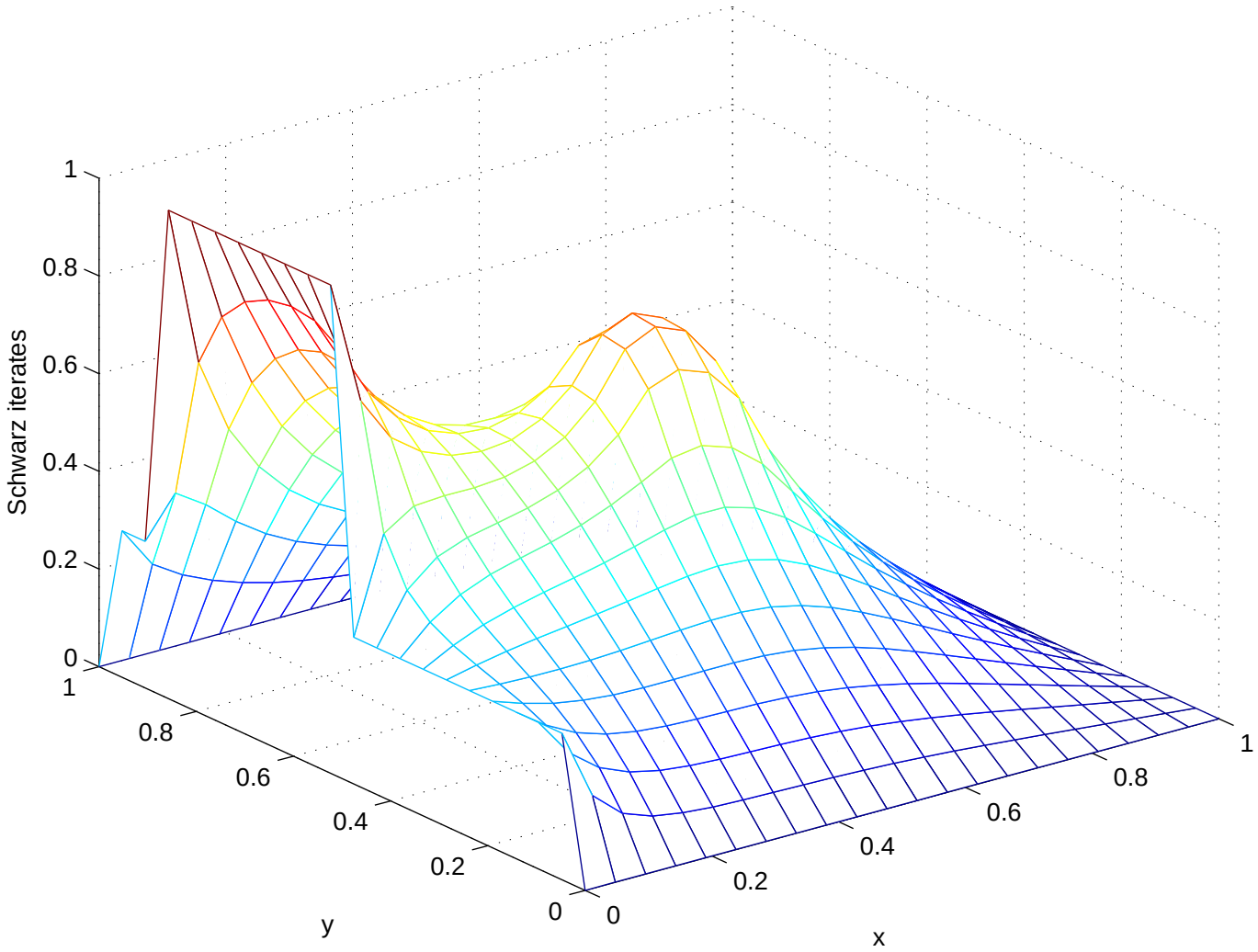
Iteration 7



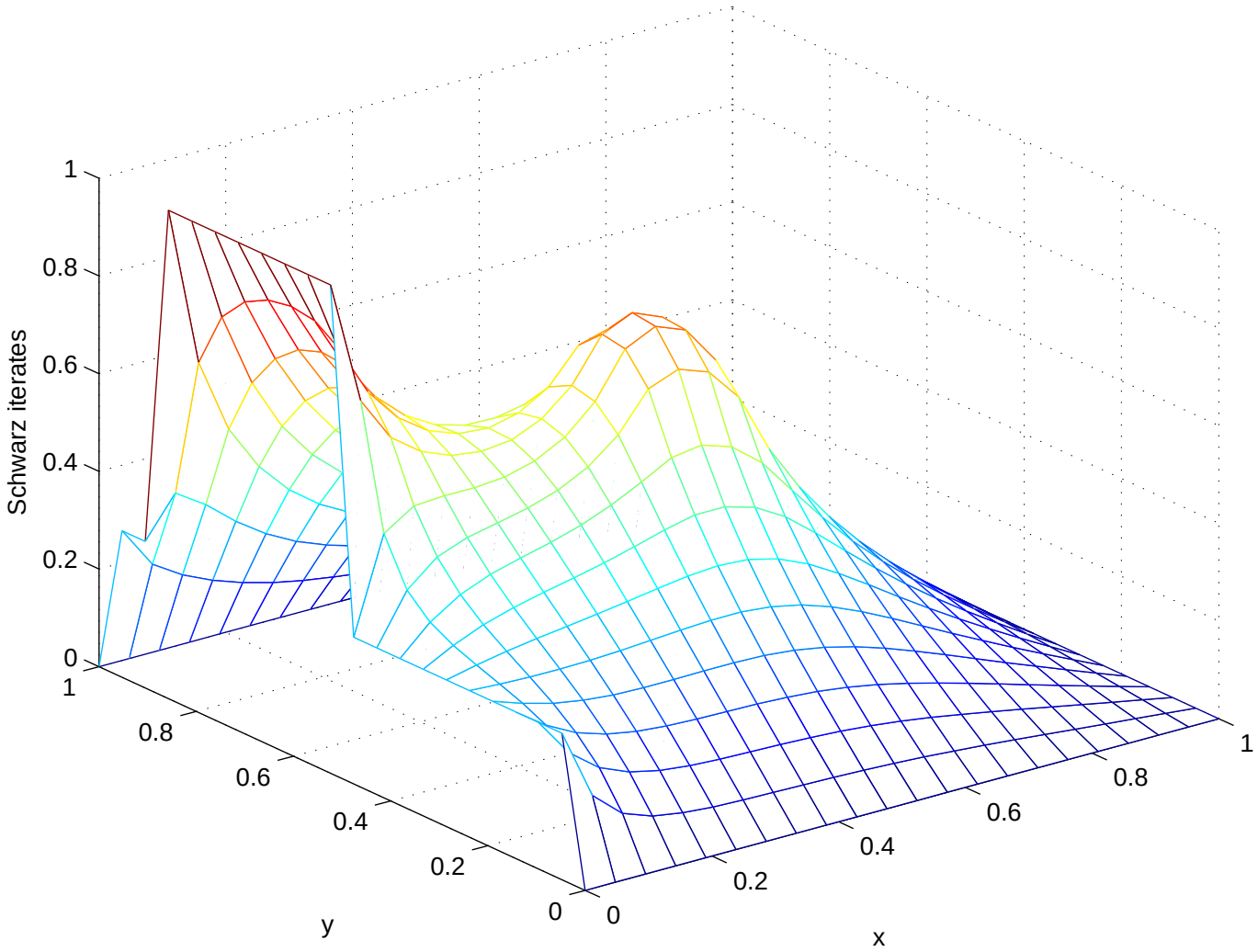
Iteration 8



Iteration 9



Iteration 10



Substructuring

Invented in the engineering community for the finite element design of aircraft (Boeing)

Przemieniecki 1963: Matrix structural analysis of substructures

The necessity for dividing a structure into substructures arises either from the requirement that different types of analysis have to be used on different components, or because the capacity of the digital computer is not adequate to cope with the analysis of the complete structure.

Idea of Przemieniecki

Przemieniecki 1963: In the present method each substructure is first analyzed separately, assuming that all common boundaries with adjacent substructures are completely fixed: these boundaries are then relaxed simultaneously and the actual boundary displacements are determined from the equations of equilibrium of forces at the boundary joints. The substructures are then analyzed separately again under the action of specified external loading and the previously determined boundary displacements.

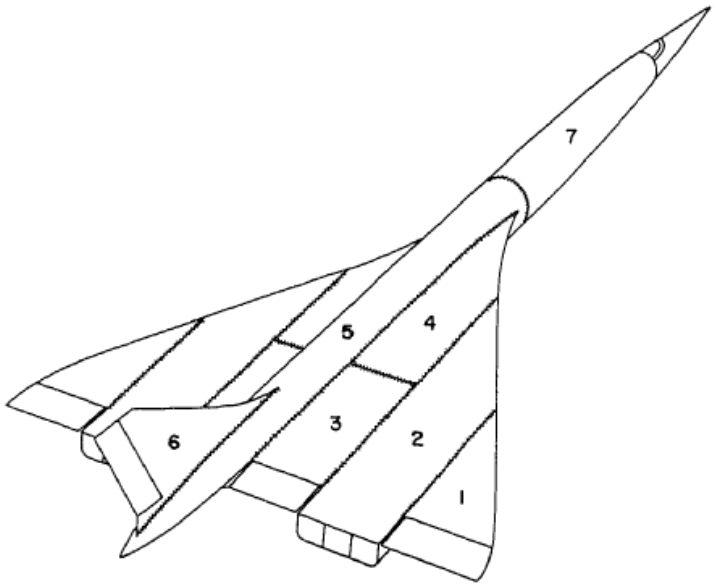


FIG. 3. Typical substructure arrangement for delta aircraft.

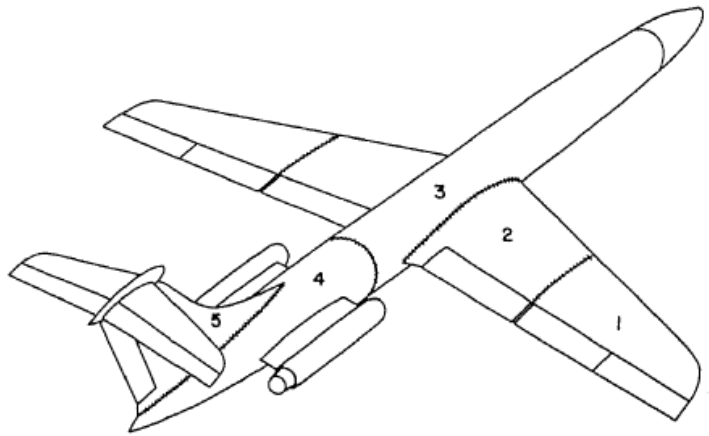


FIG. 4. Typical substructure arrangement for conventional aircraft.

Historical Example of Przemieniecki

Let P be the exterior forces, K the stiffness matrix, and U the displacement vector satisfying

$$KU = P.$$

Partition U into U_i interior in each substructure, and U_b on the interfaces between substructures:

$$K \begin{bmatrix} U_b \\ U_i \end{bmatrix} := \begin{bmatrix} K_{bb} & K_{bi} \\ K_{ib} & K_{ii} \end{bmatrix} \begin{bmatrix} U_b \\ U_i \end{bmatrix} = \begin{bmatrix} P_b \\ P_i \end{bmatrix}.$$

Physical motivation of Przemieniecki:

$$P = P^{(\alpha)} + P^{(\beta)} = \begin{bmatrix} P_b^{(\alpha)} \\ P_i \end{bmatrix} + \begin{bmatrix} P_b^{(\beta)} \\ 0 \end{bmatrix},$$

$$U = U^{(\alpha)} + U^{(\beta)} = \begin{bmatrix} 0 \\ U_i^{(\alpha)} \end{bmatrix} + \begin{bmatrix} U_b \\ U_i^{(\beta)} \end{bmatrix}.$$

Two Physically Relevant Systems

By linearity, Przemieniecki obtains two systems

$$(\alpha): \begin{bmatrix} K_{bb} & K_{bi} \\ K_{ib} & K_{ii} \end{bmatrix} \begin{bmatrix} 0 \\ U_i^{(\alpha)} \end{bmatrix} = \begin{bmatrix} P_b^{(\alpha)} \\ P_i \end{bmatrix}$$

$$(\beta): \begin{bmatrix} K_{bb} & K_{bi} \\ K_{ib} & K_{ii} \end{bmatrix} \begin{bmatrix} U_b \\ U_i^{(\beta)} \end{bmatrix} = \begin{bmatrix} P_b^{(\beta)} \\ 0 \end{bmatrix},$$

Rewriting the first one leads to

$$(\alpha): \begin{cases} K_{bi} U_i^{(\alpha)} = P_b^{(\alpha)}, \\ K_{ii} U_i^{(\alpha)} = P_i, \end{cases}$$

Knowing the forces P_i in each substructure, (α) permits to compute the interior displacements keeping interfaces fixed:

$$U_i^{(\alpha)} = K_{ii}^{-1} P_i.$$

This uncovers the unknown splitting of the interface forces

$$P_b^{(\alpha)} = K_{bi} K_{ii}^{-1} P_i,$$

Toward the Schur Complement System

Hence the remaining forces acting on the interfaces are

$$P_b^{(\beta)} := P_b - P_b^{(\alpha)} = P_b - K_{bi} K_{ii}^{-1} P_i \,,$$

One can now solve the system

$$(\beta): \begin{cases} K_{bb} U_b + K_{bi} U_i^{(\beta)} = P_b^{(\beta)}, \\ K_{ib} U_b + K_{ii} U_i^{(\beta)} = 0, \end{cases}$$

which represents the response of the structures to the interface loading $P_b^{(\beta)}$. The second equation gives the internal displacement $U_i^{(\beta)}$ based on the boundary displacement U_b,

$$U_i^{(\beta)} = -K_{ii}^{-1} K_{ib} U_b,$$

and inserting this into the first equation, Przemieniecki obtains the interface system

$$(K_{bb} - K_{bi} K_{ii}^{-1} K_{ib}) U_b = P_b - K_{bi} K_{ii}^{-1} P_i,$$

Primal Schur Method

This gives the complete interface displacement

$$U_b = (K_{bb} - K_{bi}K_{ii}^{-1}K_{ib})^{-1}(P_b - K_{bi}K_{ii}^{-1}P_i),$$

and interior displacements are obtained by summing $U_i^{(\beta)}$ and $U_i^{(\alpha)}$ (or solving $K_{ib}U_b + K_{ii}U_i = P_i$ for U_i).

Procedure of Przemieniecki:

1. Invert the block diagonal matrix K_{ii}
2. Invert the smaller matrix $S = K_{bb} - K_{bi}K_{ii}^{-1}K_{ib}$

The matrix S is called Schur complement matrix after Emilie Virginia Haynsworth (On the Schur complement 1968, Basel) after a determinant lemma of Issai Schur.

Remark: The name Schur method is more precise than substructuring, since any method can be substructured, also Schwarz methods.

Waveform Relaxation

Émile Picard (1893): Sur l'application des méthodes d'approximations successives à l'étude de certaines équations différentielles ordinaires

Les méthodes d'approximation dont nous faisons usage sont théoriquement susceptibles de s'appliquer à toute équation, mais elles ne deviennent vraiment intéressantes pour l'étude des propriétés des fonctions définies par les équations différentielles que si l'on ne reste pas dans les généralités et si l'on envisage certaines classes d'équations.

The Method of ...

14. Indiquons une autre méthode pour établir l'existence des intégrales des équations différentielles ordinaires (¹). Nous envisageons, comme plus haut, en changeant seulement un peu les notations, le système des n équations du premier ordre

$$\frac{du}{dx} = f_1(x, u, v, \ldots, w),$$

$$\frac{dv}{dx} = f_2(x, u, v, \ldots, w),$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots,$$

$$\frac{dw}{dx} = f_n(x, u, v, \ldots, w).$$

Les fonctions f sont des fonctions continues réelles des quantités réelles $x, u, v, \ldots, w$ dans le voisinage de $x_0, u_0, v_0, \ldots, w_0$. Elles sont définies quand $x, u, v, \ldots, w$ restent respectivement compris dans les intervalles

$$(x_0 - a, x_0 + a), \quad (u_0 - b, u_0 + b), \quad \ldots, \quad (w_0 - b, w_0 + b),$$

... Successive Approximations

Considérons d'abord le système

$$\frac{du_1}{dx} = f_1(x, u_0, v_0, \ldots, w_0), \qquad \ldots, \qquad \frac{dw_1}{dx} = f_n(x, u_0, v_0, \ldots, w_0),$$

nous en tirons, par quadratures, les fonctions $u_1, v_1, \ldots, w_1$, en les déterminant de manière qu'elles prennent pour x_0 les valeurs $u_0, v_0, \ldots, w_0$. On forme ensuite les équations

$$\frac{du_2}{dx} = f_1(x, u_1, v_1, \ldots, w_1), \qquad \ldots, \qquad \frac{dw_2}{dx} = f_n(x, u_1, v_1, \ldots, w_1),$$

et l'on détermine $u_2, v_2, \ldots, w_2$ par la condition qu'elles prennent respectivement pour x_0 les valeurs $u_0, v_0, \ldots, w_0$. On continue ainsi indéfiniment. Les fonctions $u_{m-1}, v_{m-1}, \ldots, w_{m-1}$ seront liées aux suivantes $u_m, v_m, \ldots, w_m$ par les relations

$$\frac{du_m}{dx} = f_1(x, u_{m-1}, v_{m-1}, \ldots, w_{m-1}),$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,$$

$$\frac{dw_m}{dx} = f_n(x, u_{m-1}, v_{m-1}, \ldots, w_{m-1}),$$

Detailed Historical Convergence Analysis

Ernest Lindelöf (1894): Sur l'application des méthodes d'approximations successives à l'étude des intégrales réelles des équations différentielles ordinaires

La présente étude a pour but de donner une exposition succincte de la méthode d'approximations successives de M. Picard en tant qu'elle s'applique aux équations différentielles ordinaires.

Theorem (Superlinear Convergence (Lindelöf 1894))

On bounded time intervals $t \in [0, T]$, the iterates satisfy the superlinear error bound

$$||\mathbf{v} - \mathbf{v}^n|| \leq \frac{(CT)^n}{n!}||\mathbf{v} - \mathbf{v}^0||,$$

where C is a positive constant.

Classical Waveform Relaxation

Lelarasmee, Ruehli and Sangiovanni-Vincentelli (1982): The Waveform Relaxation Method for Time-Domain Analysis of Large Scale Integrated Circuits.

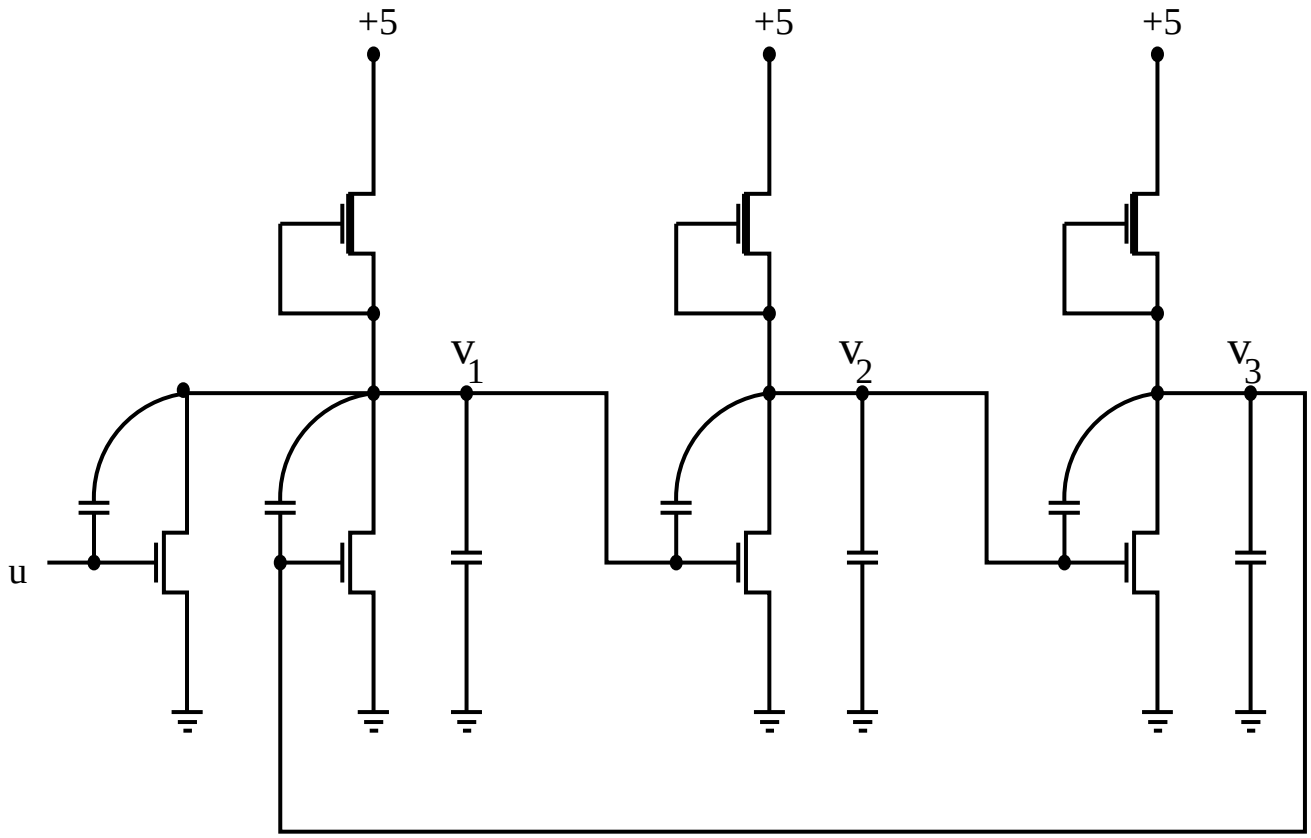
"The spectacular growth in the scale of integrated circuits being designed in the VLSI era has generated the need for new methods of circuit simulation. "Standard" circuit simulators, such as SPICE and ASTAP, simply take too much CPU time and too much storage to analyze a VLSI circuit".

Nevanlinna and Odeh (1987): Remarks on the Convergence of Waveform Relaxation Methods.

"Recently an approach called waveform relaxation methods (WR) has captured considerable attention in solving certain classes of large scale digital circuit equations."

A Historical Example

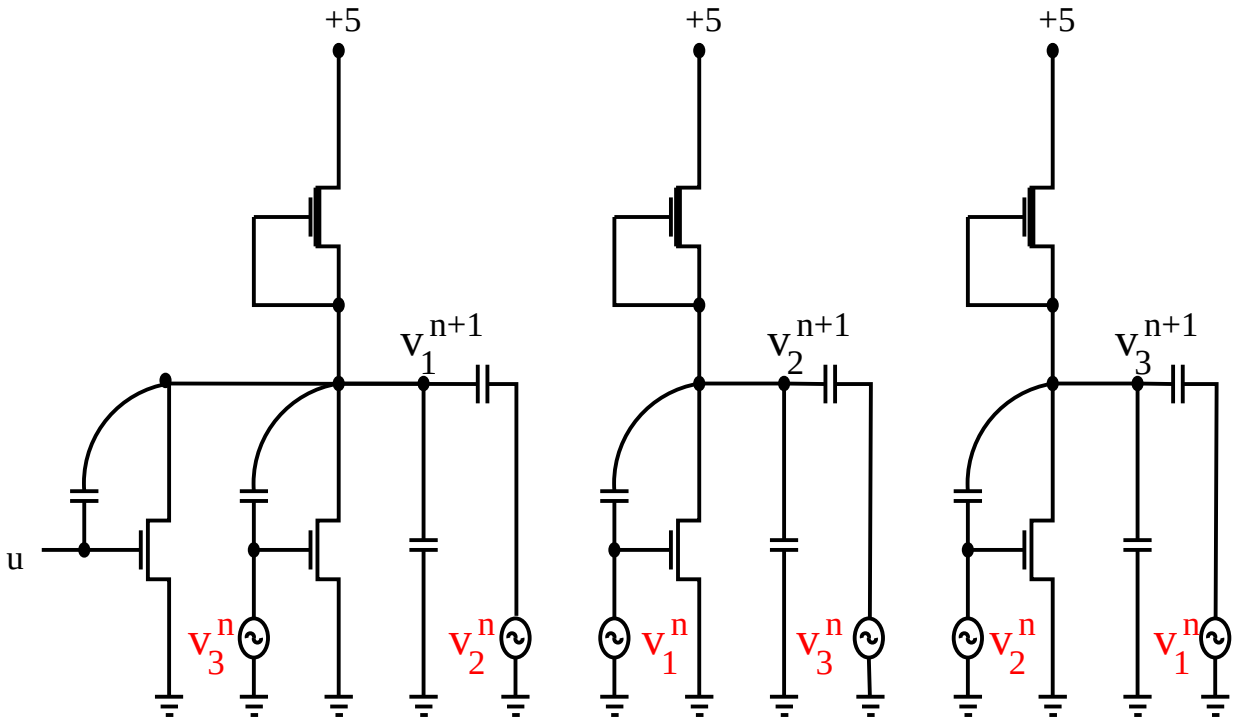
Example: a MOS ring oscillator (Lelarasmee et al 1982):



The equations for such a circuit can be written in form of a system of ordinary differential equations

$$\begin{aligned} \frac{\partial \mathbf{v}}{\partial t} &= f(\mathbf{v}), \qquad 0 < t < T \\ \mathbf{v}(0) &= \mathbf{g} \end{aligned}$$

Waveform Relaxation Decomposition

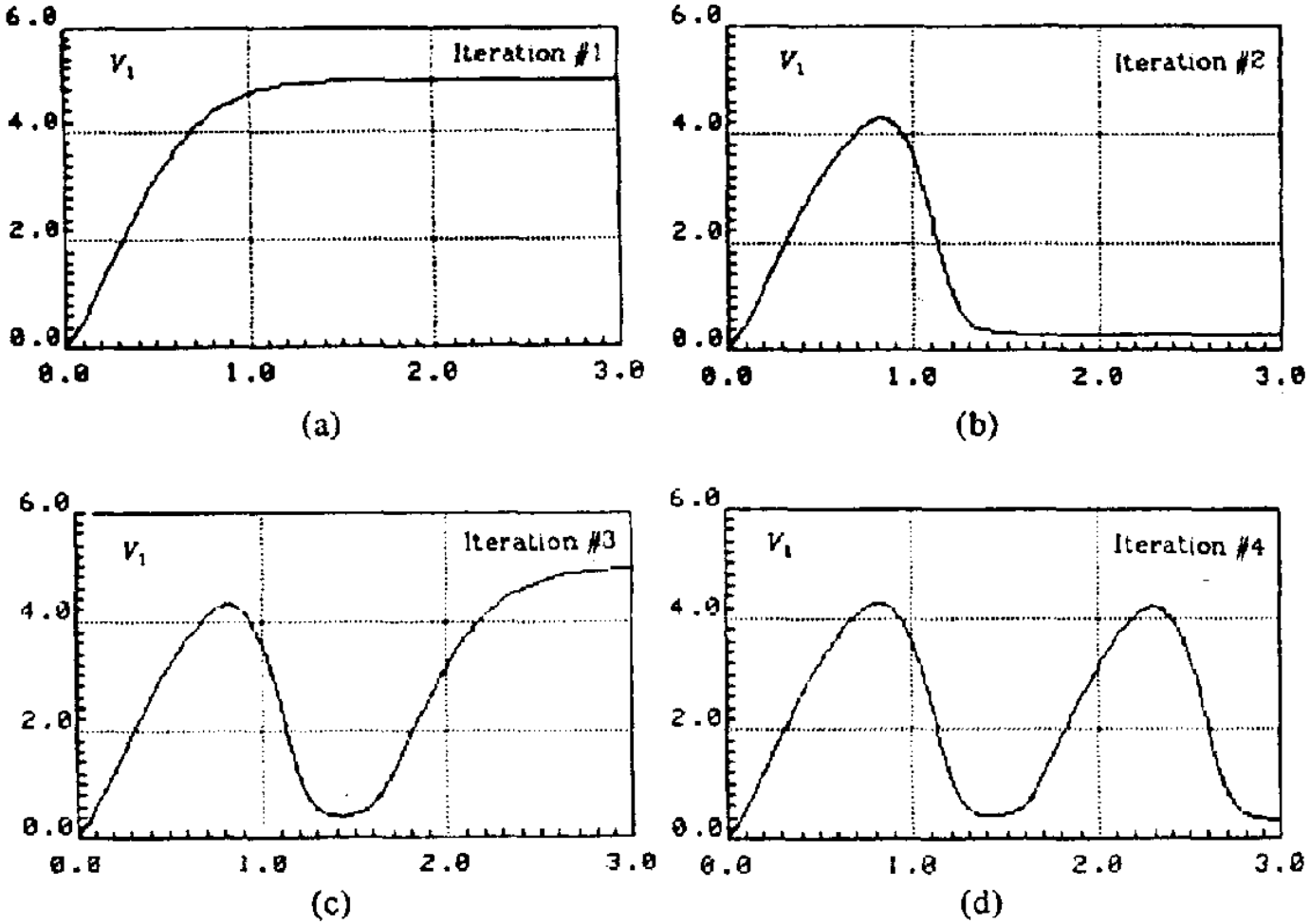


Iteration using subcircuit solutions only:

$$\begin{aligned} \partial_t v_1^{n+1} &= f_1(v_1^{n+1}, v_2^n, v_3^n) \\ \partial_t v_2^{n+1} &= f_2(v_1^n, v_2^{n+1}, v_3^n) \\ \partial_t v_3^{n+1} &= f_3(v_1^n, v_2^n, v_3^{n+1}) \end{aligned}$$

Signals along cables are called 'waveforms', which gave the algorithm its name Waveform Relaxation.

Historical Numerical Convergence Study



"Note that since the oscillator is highly nonunidirectional due to the feedback from v_3 to the NOR gate, the convergence of the iterated solutions is achieved with the number of iterations being proportional to the number of oscillating cycles of interest"

The Birth of Multigrid

Federenko (1961): A Relaxation Method for Solving Elliptic Difference Equations

The familiar iterative process [of Jacobi] is very slowly convergent. We shall try to use some special features of the convergence in order to speed it up...

We shall speak of the eigenfunctions as "good" and "bad"; the good ones include those that are smooth on the net and have few changes of sign in the domain; the bad ones often change sign and oscillate rapidly...

After a fairly small number of iterations, the error will consist of "good" eigenfunctions [...] We shall use the following method to annihilate the "good" components of the error. We introduce into the domain an auxiliary net, the step of which is q times greater than the step of the original net.

The Invention of Multigrid

Nicolaides 1975: On Multiple Grid and Related Techniques for Solving Discrete Elliptic Systems

Methods of multiple grid type: the general principle underlying this type of method was understood by pencil and paper relaxation users, and the method used by them and based on this principle was called "block relaxation."

For second-order elliptic equations in the plane [...] to reduce the error by a factor of 10^{-p} requires an amount of work proportional to the number of gridpoints n and p.

It must be said here that the implementation of a multiple grid method involves a high strategic component [...] programming a multiple grid method is a rather complex operation.

Hackbusch and Brandt

Two major competitors in the development of multigrid methods:

Brandt 1972: Multi-Level Adaptive Technique (MLAT) for Fast Numerical Solution to Boundary Value Problems

"The only disadvantage seems to be the complex programming involved"

Hackbusch 1976: A fast iterative method for solving Poisson's equation in a general region

"The basic idea – using auxiliary systems of difference equations corresponding to coarser grids – has been developed independently by the author, but it was already described by R. P. Federenko in 1961. Since then this idea has only been revived by N. S. Bakhvalov and A. Brandt."

Hackbusch and Brandt cont.

Brandt 1977: Multi-Level Adaptive Solutions to Boundary-Value Problems

"Multi-grid algorithms are not difficult to program, if the various grids are suitably organized."

"As soon as the residuals are smoothed out, convergence slows down. This is then exactly the point where relaxation sweeps should be discontinued and approximate solution of the (smoothed out) residual equations by coarser grids should be employed.".

"The basic tool is local mode (Fourier) analysis, applied to the locally linearized-frozen difference equations, ignoring boundaries."

Conclusions

- **Historical stationary iterative methods:**
  - Jacobi
  - Gauss-Seidel
  - SOR
- **Non-Stationary iterative methods:**
  - Richardson
  - Chebyshev Semi-iterative method
  - Conjugate Gradients
  - Krylov methods in general (MINRES, SYMLQ, GMRES, FOM, QMR, BiCGStab, ...)
- **Preconditioning:**
  - Comes very naturally from stationary iterative methods (Diagonal (Jacobi), Multigrid, Domain Decomposition, ...)

Always use Krylov acceleration!