Fachbereich Mathematik
Prof. Dr. P. Spellucci
Sommersemester 2009

TECHNISCHE UNIVERSITÄT DARMSTADT

6.7.2009

Numerische Lineare Algebra
Übung 11

Präsenzübung

Ü 32 Lösen Sie das Gleichungssystem $Ax = b$ mit:

$$A = \begin{pmatrix} 5 & 1 & 1 \\ 1 & 5 & -1 \\ 1 & -1 & 5 \end{pmatrix} \quad b = \begin{pmatrix} 4 \\ 2 \\ -4 \end{pmatrix}$$

mit dem cg-Verfahren von Hestenes und Stiefel und dem Startvektor $x_0 = 0$. Rechnen Sie exakt mit Brüchen.

Ü 33 Sei $A \in \mathbb{R}^{n,n}$ symmetrisch und positiv definit mit $k < n$ verschiedenen (reellen) Eigenwerten $\lambda_1, ..., \lambda_k$. Man zeige, daß in diesem Fall das cg–Verfahren schon nach k Iterationen die Lösung von $Ax^* = b$ liefert, d.h. $x_k = x^*$.
Hinweis: Nach Satz 2.4.3 gilt mit

$$E(x) \;=\; \tfrac{1}{2}(x - x^*)^T A(x - x^*) \;=\; f(x) + \tfrac{1}{2}(x^*)^T Ax^*$$

dass

$$E(x_k) \;\leq\; E(x_0) \min_{p_k} \max_{i} |p_k(\lambda_i)| \quad \text{wo } p_k \;\in\; \Pi_k \;:\; p_k(0) \;=\; 1\;.$$

Ü 34 Über das cg-Verfahren von Stiefel und Hestenes wurde in der Vorlesung Folgendes bewiesen: Es sei mit

$$r \;=\; Ax - b$$

$$\begin{aligned} p_0 &= r_0 \\ x_{i+1} &= x_i - \sigma_i p_i \;,\; r_{i+1}^T p_i \;=\; 0 \text{ definiert } \sigma_i \\ r_{i+1} &= r_i - \sigma_i A p_i \\ p_{i+1} &= r_{i+1} + \frac{\| r_{i+1} \|^2}{\| r_i \|^2} p_i \end{aligned}$$

Dann gilt

$$\begin{aligned} p_i^T A p_j &= 0 \quad \text{für} \quad i \neq j \\ r_k^T &\perp \operatorname{span}(p_0, \ldots, p_{k-1}) \end{aligned}$$

mit $i,\; j \;\leq\; N \;\leq\; n$.
Zeigen Sie

$$r_j^T r_i \;=\; 0 \text{ für } i \;\neq\; j \;,\; i,j \;\leq\; N\;.$$

Hausübung

H 32 Seien $a >> 1$, $b = (0,0)^T$, $A = \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix}$. Gesucht werde das Minimum von

$$f(x) \;=\; \tfrac{1}{2} x^T A x - b^T x \;.$$

Dazu werde das Gradientenverfahren mit dem Startvektor $x_0 = (a,1)^T$ angewendet, d.h.

$$x_{k+1} \;=\; x_k - \sigma_k \nabla f(x_k), \quad \sigma_k \;=\; \frac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_k)^T A \nabla f(x_k)} \;.$$

- Zeigen Sie, daß die Abstiegsrichtungen r_k und r_{k+1} aufeinanderfolgender Iterationsschritte aufeinander senkrecht stehen. Was folgt daraus für zweidimensionale Probleme?
- Zeigen Sie, daß in diesem Beispiel für die durch das Gradientenverfahren gelieferte Folge mit den Gliedern $x_k = (\xi_k, \eta_k)^T$

  $$\xi_{k+1} = \rho \xi_k, \quad \eta_{k+1} = -\rho \eta_k, \;\; k = 0,1,\ldots$$

  mit $\rho = \frac{a-1}{a+1}$ gilt. Was bedeutet dieses Ergebnis für die Konvergenzrate?
- Zeigen Sie, daß in diesem Beispiel r_k und r_{k+1} bezüglich der durch A induzierten Metrik fast parallel sind.

H 33 Man bestimme die Minima von

a)

$$f(x) \quad = \quad \tfrac{1}{100}(10^4 \xi_1^2 + 3000 \xi_1 \xi_2 + 9750 \xi_2^2 + \xi_1 \xi_3 + \xi_3^2 + 300 \xi_1 - 400 \xi_2 + \xi_3)$$

b)

$$f(x) \quad = \quad \xi_1^2 + 0.3 \xi_1 \xi_2 + 0.975 \xi_2^2 + 0.01 \xi_1 \xi_3 + \xi_3^2 + 3 \xi_1 - 4 \xi_2 + \xi_3$$

mit dem cg-Verfahren von Hestenes und Stiefel. Man trage $-\lg(||x_k - x^*|| / ||x^*||)$ als Funktion von k auf. Starten Sie mit $x_0 = 0$. Im Fall a) ist:

$$x^* = (-0.018478819, 0.023355715, -0.49076058)^T$$

Welcher Zusammenhang besteht zwischen diesen beiden Aufgaben?

H 34 PCG: Wegen Satz 2.4.3 ist die Konvergenz des CG-Verfahrens für $Ax = b$ langsam, wenn die Eigenwerte von A weit auseinander liegen. Deshalb versucht man durch eine Kongruenztransformation, die Verteilung der Eigenwerte zu verbessern. Dieses Vorgehen wird Präkonditionieren genannt und führt zu einem modifizierten Algorithmus, dem Verfahren PCG.

Sei $M = LL^T$ mit der regulären Matrix L.
1.) Zeigen Sie, daß M symmetrisch und positiv definit ist.

Statt der Aufgabe $Ax = b$ wird die äquivalente Form $L^{-1}AL^{-T}L^Tx = L^{-1}b$ herangezogen. Zur Abkürzung wird

$$\hat{A} = L^{-1}AL^{-T}\,,\ \hat{x} = L^Tx \quad \text{und} \quad \hat{b} = L^{-1}b \qquad (*)$$

gesetzt.
2.) Zeigen Sie, daß $\hat{A}$ symmetrisch, positiv definit und ähnlich zu $M^{-1}A$ ist.
Bem.: Dies gibt Hinweise, wie die Matrix M und damit L, L^T zu wählen sind.
3.) Wie lautet der CG-Algorithmus für die Aufgabe $\hat{A}\hat{x} = \hat{b}$?

Aus Aufwandsgründen möchte man zwar das transformierte Gleichungssystem lösen, aber gleichzeitig vermeiden, die Matrix $\hat{A}$ explizit bestimmen zu müssen.

Im Folgenden sollen mit $\hat{r}_k, \hat{p}, ...$ etc. die Zwischenwerte des CG-Verfahrens angewandt auf $\hat{A}\hat{x} = \hat{b}$ bezeichnen. Für den modifizierten Algorithmus definiere wegen (*) als Zwischenwerte $s_k = L\hat{r}_k$, $y_k = L^{-T}\hat{x}_k$ und $q_k = L\hat{p}_k$. Diese Schreibweise soll andeuten, daß s_k, y_k und q_k nur indirekt mit den entsprechenden Zwischenergebnissen des ursprünglichen CG-Verfahrens angewandt auf $Ax = b$ zu tun haben.
4.) Zeigen Sie

$$\begin{aligned} y_k &= y_{k-1} + \hat{\alpha}_k M^{-1} q_{k-1} \\ s_k &= s_{k-1} - \hat{\alpha}_k A M^{-1} q_{k-1} \\ M^{-1} q_k &= M^{-1} s_k + \hat{\beta}_k M^{-1} q_{k-1}\,. \end{aligned}$$

Zur besseren Übersichtlichkeit werden die Hilfsvektoren $g_k = M^{-1}q_k$ und $z = M^{-1}s_k$ eingeführt.
5.) Zeigen Sie $\hat{r}_k^T\hat{r}_k = s_k^T z$ und $\hat{p}_k^T\hat{A}\hat{p}_k = g_k^T A g_k$.
6.) Zeigen Sie, daß der im Skript angegebene Algorithmus für das präkonditionierte CG–Verfahren die gewünschten Anforderungen erfüllt. Bem.: Da Gleichungssysteme der Form $Mz = s$ gelöst werden müssen, liefert dies weitere Richtlinien für eine Wahl von M bzw. L, L^T.

Numerische Lineare Algebra
Übung 11, Lösungsvorschlag

Präsenzübung

Ü 32 Lösen Sie das Gleichungssystem $Ax = b$ mit:

$$A = \begin{pmatrix} 5 & 1 & 1 \\ 1 & 5 & -1 \\ 1 & -1 & 5 \end{pmatrix} \quad b = \begin{pmatrix} 4 \\ 2 \\ -4 \end{pmatrix}$$

mit dem cg-Verfahren von Hestenes und Stiefel und dem Startvektor $x_0 = 0$. Rechnen Sie exakt mit Brüchen.

Mit $r(x) := Ax - b$ erhalten wir:

$$\begin{array}{rclrclrcl} r(x_0) & = & (-4,-2,4)^T & p_0 & = & (-4,-2,4)^T & (r(x_0))^T p_0 & = & 36 \\ Ap_0 & = & 18(-1,-1,1)^T & p_0^T A p_0 & = & 180 & \sigma_0 & = & 1/5 \end{array}$$

$$x_1 = \frac{2}{5}(2,1,-2)^T$$

$$\begin{array}{rclrclrcl} r(x_1) & = & \frac{2}{5}(-1,4,1)^T & p_1 & = & \frac{18}{25}(-1,2,1)^T & (r(x_1))^T p_1 & = & 4\frac{18}{25} \\ Ap_1 & = & \frac{18}{25}(-2,8,2)^T & p_1^T A p_1 & = & 20\left(\frac{18}{25}\right)^2 & \sigma_1 & = & \frac{5}{18} \end{array}$$

$$x_2 = (1,0,-1)^T$$

Ü 33 Sei $A \in \mathbb{R}^{n,n}$ symmetrisch und positiv definit mit $k < n$ verschiedenen (reellen) Eigenwerten $\lambda_1, ..., \lambda_k$. Man zeige, daß in diesem Fall das cg–Verfahren schon nach k Iterationen die Lösung von $Ax^* = b$ liefert, d.h. $x_k = x^*$.
Hinweis: Nach Satz 2.4.3 gilt mit

$$E(x) = \tfrac{1}{2}(x - x^*)^T A (x - x^*) = f(x) + \tfrac{1}{2}(x^*)^T A x^*$$

dass

$$E(x_k) \leq E(x_0) \min_{p_k} \max_i |p_k(\lambda_i)| \quad \text{wo } p_k \in \Pi_k : p_k(0) = 1 .$$

Für den Fall $\lambda_k = ... = \lambda_n$ liefert diese Aussage die Behauptung. Denn man wähle dazu in der Aussage

$$P_k(\lambda) = \left(\prod_{i=1}^{k} \frac{\lambda_i - \lambda}{\lambda_i}\right)$$

Ü 34 Über das cg-Verfahren von Stiefel und Hestenes wurde in der Vorlesung Folgendes bewiesen: Es sei mit

$$r = Ax - b$$

$$\begin{aligned}
p_0 &= r_0 \\
x_{i+1} &= x_i - \sigma_i p_i \ , \ r_{i+1}^T p_i = 0 \text{ definiert } \sigma_i \\
r_{i+1} &= r_i - \sigma_i A p_i \\
p_{i+1} &= r_{i+1} + \frac{\| r_{i+1} \|^2}{\| r_i \|^2} p_i
\end{aligned}$$

Dann gilt

$$\begin{aligned}
p_i^T A p_j &= 0 \quad \text{für} \quad i \neq j \\
r_k^T &\perp \text{span}(p_0, \ldots, p_{k-1})
\end{aligned}$$

mit $i, \ j \leq N \leq n$.
Zeigen Sie

$$r_j^T r_i = 0 \text{ für } i \neq j \ , \ i, j \leq N \ .$$

Zu zeigen ist lediglich

$$span(p_0, \ldots, p_k) = span(r_0, \ldots, r_k)$$

Wegen $p_0 = r_0$ ist die Induktionsverankerung gegeben. Induktionsschritt:

$$\begin{aligned}
span(p_0, \ldots, p_k) &= span(p_0, \ldots, p_{k-1}) \oplus span(p_k) \\
&= span(r_0, \ldots, r_{k-1}) \oplus span(r_k + \beta_k p_{k-1}) \\
&= span(r_0, \ldots, r_{k-1}) \oplus span(r_k) \oplus span(p_{k-1}) \\
&= span(r_0, \ldots, r_k)
\end{aligned}$$

Hausübung

H 32 Seien $a >> 1$, $b = (0,0)^T$, $A = \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix}$. Gesucht werde das Minimum von

$$f(x) \;=\; \tfrac{1}{2} x^T A x - b^T x \;.$$

Dazu werde das Gradientenverfahren mit dem Startvektor $x_0 = (a,1)^T$ angewendet, d.h.

$$x_{k+1} \;=\; x_k - \sigma_k \nabla f(x_k), \quad \sigma_k \;=\; \frac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_k)^T A \nabla f(x_k)} \;.$$

- Zeigen Sie, daß die Abstiegsrichtungen r_k und r_{k+1} aufeinanderfolgender Iterationsschritte aufeinander senkrecht stehen. Was folgt daraus für zweidimensionale Probleme?
- Zeigen Sie, daß in diesem Beispiel für die durch das Gradientenverfahren gelieferte Folge mit den Gliedern $x_k = (\xi_k, \eta_k)^T$

  $$\xi_{k+1} = \rho \xi_k, \quad \eta_{k+1} = -\rho \eta_k, \; k = 0,1,\ldots$$

  mit $\rho = \frac{a-1}{a+1}$ gilt. Was bedeutet dieses Ergebnis für die Konvergenzrate?
- Zeigen Sie, daß in diesem Beispiel r_k und r_{k+1} bezüglich der durch A induzierten Metrik fast parallel sind.

- *Es gelten* $r_k = \nabla f(x_k) = A x_k - b$, $x_{k+1} = x_k - \alpha r_k$, *wobei* α *den Wert von*

  $$f(x_k - \alpha r_k) = \frac{1}{2} x_k^T A x_k - b^T x_k - \alpha x_k^T A r_k + \frac{1}{2} \alpha^2 r_k^T A r_k + \alpha b^T r_k$$

  minimieren soll, daraus folgt

  $$-x_k^T A r_k + b^T r_k + \alpha r_k^T A r_k = 0 \text{ bzw. } \alpha = \frac{r_k^T r_k}{r_k^T A r_k}.$$

  Damit erhält man $r_k^T r_{k+1} = r_k^T (A x_{k+1} - b) = -r_k^T b + r_k^T A x_k - \alpha r_k^T A r_k = 0$. *Im* $\mathbb{R}^2$ *sind also* r_k *und* r_{k+2} *parallel. In höheren Dimensionen hat man diese Zyklenbildung im allgemeinen nicht.*

- *Es gilt* $x_{k+1} = x_k - \alpha r_k = \begin{pmatrix} \xi_k \\ \eta_k \end{pmatrix} - \frac{\xi_k^2 + a^2 \eta_k^2}{\xi_k^2 + a^3 \eta_k^2} \begin{pmatrix} \xi_k \\ a \eta_k \end{pmatrix}$. *Zu zeigen sind* $\xi_k = \rho^k a$, $\eta_k = (-\rho)^k$, $k = 0,1,\ldots$ *Für* $k = 0$ *ist dies nach Voraussetzung erfüllt. Induktiv folgt*

  $$\begin{pmatrix} \xi_{k+1} \\ \eta_{k+1} \end{pmatrix} = \begin{pmatrix} \rho^k a \\ (-\rho)^k \end{pmatrix} - \frac{2\rho^{2k} a^2}{\rho^{2k}(a^2 + a^3)} \begin{pmatrix} \rho^k a \\ (-\rho)^k a \end{pmatrix} = \begin{pmatrix} \rho^{k+1} a \\ (-\rho)^{k+1} \end{pmatrix}.$$

  Die Konvergenzrate ist damit gleich ρ *und deshalb sehr nahe bei 1.*

- *Für den Winkel φ zwischen r_k und r_{k+1} gilt*

$$\cos\varphi = \frac{r_{k+1}^T A r_k}{\sqrt{r_k^T A r_k}\sqrt{r_{k+1}^T A r_{k+1}}} = \frac{a^2\rho^{2k+1}(1-a)}{\sqrt{a^2\rho^{2k}(1+a)}\sqrt{a^2\rho^{2k+2}(1+a)}} = \rho.$$

Daraus folgt $\varphi \approx 0$.

H 33 Man bestimme die Minima von

a)

$$f(x) \quad = \quad \tfrac{1}{100}(10^4\xi_1^2 + 3000\xi_1\xi_2 + 9750\xi_2^2 + \xi_1\xi_3 + \xi_3^2 + 300\xi_1 - 400\xi_2 + \xi_3)$$

b)

$$f(x) \quad = \quad \xi_1^2 + 0.3\xi_1\xi_2 + 0.975\xi_2^2 + 0.01\xi_1\xi_3 + \xi_3^2 + 3\xi_1 - 4\xi_2 + \xi_3$$

mit dem cg-Verfahren von Hestenes und Stiefel. Man trage $-\lg(||x_k - x^*||/||x^*||)$ als Funktion von k auf. Starten Sie mit $x_0 = 0$. Im Fall a) ist:

$$x^* = (-0.018478819, 0.023355715, -0.49076058)^T$$

Welcher Zusammenhang besteht zwischen diesen beiden Aufgaben?

Mit 8–stelliger Rechnung ergab sich:

a)

$$\nabla f(x) \quad = \quad \frac{1}{100}\begin{pmatrix} 20000x_1 & + & 3000x_2 & + & x_3 & + & 300 \\ 3000x_1 & + & 19500x_2 & & & - & 400 \\ x_1 & & & + & 2x_3 & + & 1 \end{pmatrix}$$

$$||\nabla f(x_0)||_2 = 25.000098 \quad \sigma_0 = 5.9524041E\text{-}3$$

$$\begin{aligned} p_0 &= (3, -4, .01)^T \\ x_1 &= (-0.017857213, 0.023809616, -5.9524041E\text{-}5)^T \end{aligned}$$

$$||\nabla f(x_1)||_2 = .031984221 \quad \sigma_1 = 4.4227653E\text{-}3$$

$$\begin{aligned} p_1 &= (0.14668378, 0.10204121, -9.8330307E\text{-})^T 3 \\ x_2 &= (-1.8505957, 2.3358312E\text{-}2, -1.0301323E\text{-}4)^T \end{aligned}$$

$$||\nabla f(x_2)||_2 = 9.658395E\text{-}5 \quad \sigma_2 = 49.85020$$

$$p_2 = (-2.745139E\text{-}7, 1.0086034E\text{-}7, -9.842573E\text{-}3)^T$$
$$x_3 = (-1.8492273E\text{-}2, 2.3353283E\text{-}2, -0.49075731)^T$$

$$||\nabla f(x_3)||_2 \quad = \quad 8.4102547E\text{-}6$$

b) *Man erhält das Problem b) aus dem Problem a) durch die Substitution*

$$\xi_1 \to \xi_1/100 \quad \xi_2 \to \xi_2/100$$

mit folgender Lösung:

$$x^* = (-1.8478819, 2.335571, -4.90765056E\text{-}1)^T$$

Die Konditionszahl verbessert sich dadurch extrem. Man sollte also in diesem Zusammenhang eine positiv definite Matrix stets so umskalieren, dass ihre Diagonale ein Vielfaches der Einheitsmatrix wird:

$$A \;\to\; D^{-1/2}AD^{-1/2} \quad \mathit{mit} \quad D \;=\; \mathit{diag}(A)\ .$$

$$\nabla f(x) \;=\; \begin{pmatrix} 2x_1 & + & 0.3x_2 & + & 0.01x_3 & + & 3 \\ 0.3x_1 & + & 1.95x_2 & & & - & 4 \\ 0.01x_1 & & & + & 2x_3 & + & 1 \end{pmatrix}$$

$$||\nabla f(x_0)||_2 = 26 \quad \sigma_0 = 0.5901044$$

$$\begin{aligned} p_0 &= (3, -4, 1)^T \\ x_1 &= (-1.7703132, 2.360418, -0.5901044)^T \end{aligned}$$

$$||\nabla f(x_1)||_2 = 7.0426821E\text{-}2 \quad \sigma_1 = 0.48044234$$

$$\begin{aligned} p_1 &= (0.16972389, 0.060885703, -0.19520325)^T \\ x_2 &= (-1.8518558, 2.331165, -0.49632048)^T \end{aligned}$$

$$||\nabla f(x_2)||_2 = 3.0721312E\text{-}4 \quad \sigma_2 = 0.46289879$$

$$\begin{aligned} p_2 &= (-8.5846644E\text{-}3, -9.5181512E\text{-}3, -1.2011084E\text{-}2)^T \\ x_3 &= (-1.8478819, 2.335571, -0.49076062)^T \end{aligned}$$

H 34 PCG: Wegen Satz 2.4.3 ist die Konvergenz des CG-Verfahrens für $Ax = b$ langsam, wenn die Eigenwerte von A weit auseinander liegen. Deshalb versucht man durch eine Kongruenztransformation, die Verteilung der Eigenwerte zu verbessern. Dieses Vorgehen wird Präkonditionieren genannt und führt zu einem modifizierten Algorithmus, dem Verfahren PCG.

Sei $M = LL^T$ mit der regulären Matrix L.
1.) Zeigen Sie, daß M symmetrisch und positiv definit ist.

Statt der Aufgabe $Ax = b$ wird die äquivalente Form $L^{-1}AL^{-T}L^Tx = L^{-1}b$ herangezogen. Zur Abkürzung wird

$$\hat{A} = L^{-1}AL^{-T}\,,\ \hat{x} = L^Tx \quad \text{und} \quad \hat{b} = L^{-1}b \qquad (*)$$

gesetzt.
2.) Zeigen Sie, daß $\hat{A}$ symmetrisch, positiv definit und ähnlich zu $M^{-1}A$ ist.
Bem.: Dies gibt Hinweise, wie die Matrix M und damit L, L^T zu wählen sind.
3.) Wie lautet der CG-Algorithmus für die Aufgabe $\hat{A}\hat{x} = \hat{b}$?

Aus Aufwandsgründen möchte man zwar das transformierte Gleichungssystem lösen, aber gleichzeitig vermeiden, die Matrix $\hat{A}$ explizit bestimmen zu müssen.

Im Folgenden sollen mit $\hat{r}_k, \hat{p}, ...$ etc. die Zwischenwerte des CG-Verfahrens angewandt auf $\hat{A}\hat{x} = \hat{b}$ bezeichnen. Für den modifizierten Algorithmus definiere wegen (*) als Zwischenwerte $s_k = L\hat{r}_k$, $y_k = L^{-T}\hat{x}_k$ und $q_k = L\hat{p}_k$. Diese Schreibweise soll andeuten, daß s_k, y_k und q_k nur indirekt mit den entsprechenden Zwischenergebnissen des ursprünglichen CG-Verfahrens angewandt auf $Ax = b$ zu tun haben.
4.) Zeigen Sie

$$
\begin{aligned}
y_k &= y_{k-1} + \hat{\alpha}_k M^{-1} q_{k-1} \\
s_k &= s_{k-1} - \hat{\alpha}_k A M^{-1} q_{k-1} \\
M^{-1} q_k &= M^{-1} s_k + \hat{\beta}_k M^{-1} q_{k-1} \,.
\end{aligned}
$$

Zur besseren Übersichtlichkeit werden die Hilfsvektoren $g_k = M^{-1}q_k$ und $z = M^{-1}s_k$ eingeführt.
5.) Zeigen Sie $\hat{r}_k^T \hat{r}_k = s_k^T z$ und $\hat{p}_k^T \hat{A} \hat{p}_k = g_k^T A g_k$.
6.) Zeigen Sie, daß der im Skript angegebene Algorithmus für das präkonditionierte CG–Verfahren die gewünschten Anforderungen erfüllt. Bem.: Da Gleichungssysteme der Form $Mz = s$ gelöst werden müssen, liefert dies weitere Richtlinien für eine Wahl von M bzw. L, L^T.

Der Algorithmus bekommt in der Schreibweise, die in der Aufgabenstellung eingeführt wurde die Form:

$$
\begin{aligned}
s_0 &= b - Ax_0 \\
M z_0 &= s_0 \,,\ g_0 = z_0 && \textit{Gleichungssystem lösen} \\
k = 1, 2 ... & \\
\hat{\alpha}_k &= s_{k-1}^T z_{k-1} / (g_{k-1}^T A g_{k-1}) \\
y_k &= y_{k-1} + \hat{\alpha}_k g_{k-1} \\
s_k &= s_{k-1} - \hat{\alpha}_k A g_{k-1} \\
M z_k &= s_k && \textit{Gleichungssystem lösen} \\
\hat{\beta}_k &= s_k^T z_k / (s_{k-1}^T z_{k-1}) \\
g_k &= z_k + \hat{\beta}_k g_{k-1}
\end{aligned}
$$