ИВМ СО РАН
Отдел вычислительной математики

Киреев И. В.

МЕТОД СОПРЯЖЁННЫХ ГРАДИЕНТОВ

Красноярск 2011

1. Постановка задачи

$$B\boldsymbol{x} = \boldsymbol{b}; \qquad \boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^n. \tag{1}$$

$$B^T = B, \qquad 0 < \lambda_1 \le \lambda_B \le \lambda_n. \tag{2}$$

Тогда

$$\boldsymbol{x} = \operatorname*{argmin}_{\boldsymbol{y} \in \mathbb{R}^n} \varphi(\boldsymbol{y}),$$

где

$$\varphi(\boldsymbol{y}) = \frac{1}{2}\langle B\boldsymbol{y}, \boldsymbol{y}\rangle - \langle \boldsymbol{b}, \boldsymbol{y}\rangle,\ \boldsymbol{y} \in \mathbb{R}^n, \tag{3}$$

или

$$\varphi(\boldsymbol{y}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_B^2 - \frac{1}{2}\|\boldsymbol{x}\|_B^2,\ \|\boldsymbol{y}\|_B^2 = \langle B\boldsymbol{y}, \boldsymbol{y}\rangle.$$

2. **Теорема 1** (Об относительной погрешности решения линейной системы).

$$(B + \mathcal{E}_B)(\boldsymbol{x} + \varepsilon_{\boldsymbol{x}}) = \boldsymbol{b} + \varepsilon_{\boldsymbol{b}}.$$

- возмущенная система. Тогда

$$\frac{\|\varepsilon_{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|} \leq \frac{æ(B)}{1 - æ(B) \cdot \frac{\|\mathcal{E}_B\|}{\|B\|}} \left(\frac{\|\mathcal{E}_B\|}{\|B\|} + \frac{\|\varepsilon_{\boldsymbol{b}}\|}{\|\boldsymbol{b}\|} \right),$$

где $\|B\|$-спектральная норма матрицы B :

$$\|B\| = \sup_{\boldsymbol{y} \neq \boldsymbol{0}} \frac{\|B\boldsymbol{y}\|}{\|\boldsymbol{y}\|} = \lambda_n, æ(B) = \|B\| \cdot \|B^{-1}\| = \frac{\lambda_n}{\lambda_1};$$

3. **Общая схема методов спуска**

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha_k \boldsymbol{s}^k, \tag{4}$$

$$\boldsymbol{r}^k = B\boldsymbol{x}^k - \boldsymbol{b} = \mathrm{grad}\varphi(\boldsymbol{y})\big|_{\boldsymbol{y}=\boldsymbol{x}^k}. \tag{5}$$

$$\alpha_k = \underset{\alpha\in\mathbb{R}}{\mathrm{argmin}}\,\varphi(\boldsymbol{x}^k + \alpha\boldsymbol{s}^k) = -\frac{\langle \boldsymbol{r}^k, \boldsymbol{s}^k\rangle}{\langle B\boldsymbol{s}^k, \boldsymbol{s}^k\rangle},$$

$$\varphi(\boldsymbol{x}^{k+1}) = \varphi(\boldsymbol{x}^k) - \frac{1}{2}\frac{\langle \boldsymbol{r}^k, \boldsymbol{s}^k\rangle^2}{\langle B\boldsymbol{s}^k, \boldsymbol{s}^k\rangle}.$$

4. **Теорема 2** (О достаточном условии сходимости релаксационного процесса). *Пусть для процесса* (4) *выполнено ограничение*

$$|\cos\angle(\boldsymbol{r}^k, \boldsymbol{s}^k)| = \frac{|\langle \boldsymbol{r}^k, \boldsymbol{s}^k\rangle|}{\|\boldsymbol{r}^k\|\|\boldsymbol{s}^k\|} \geq C_s > 0$$

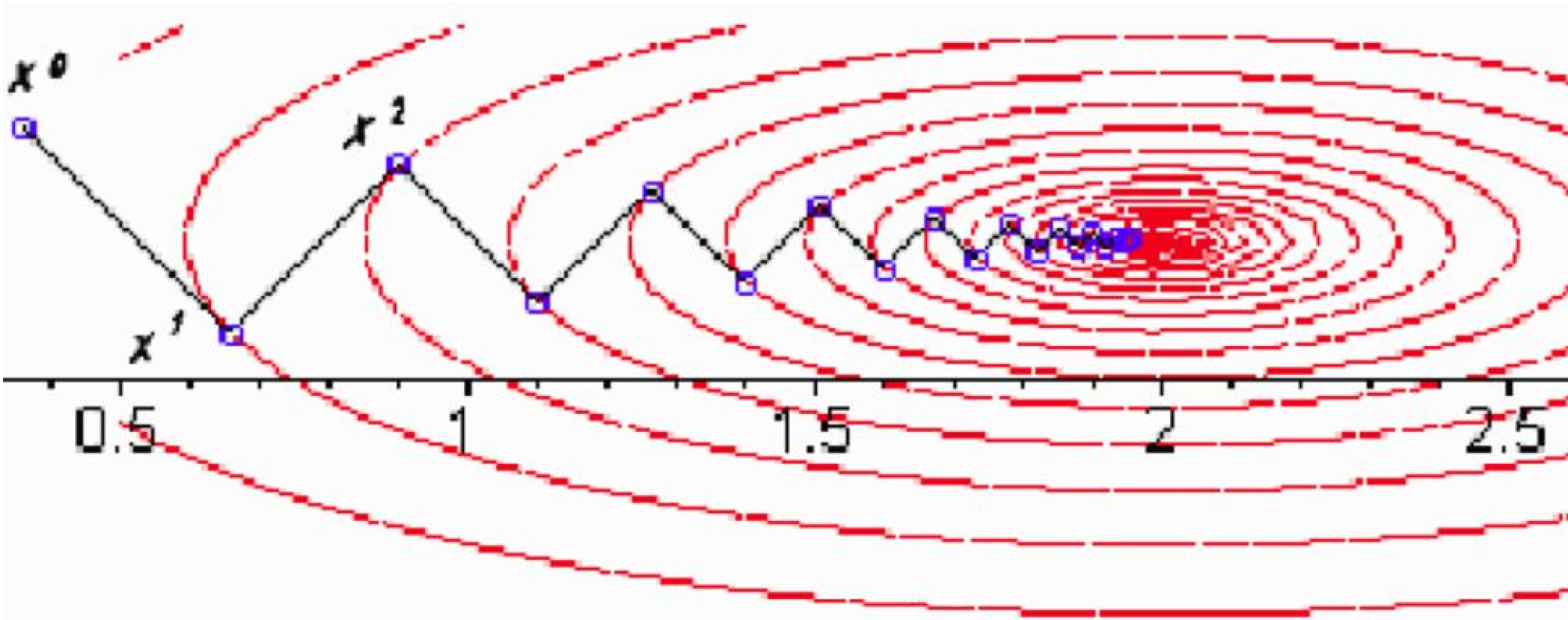
Тогда $\boldsymbol{x}^k \xrightarrow[k\to\infty]{} \boldsymbol{x}$ *при любом* $\boldsymbol{x}^0$ *и*

$$\|\boldsymbol{x}^k - \boldsymbol{x}\|_B \leq \mu^k \|\boldsymbol{x}^0 - \boldsymbol{x}\|_B,$$

где

$$\mu \leq \sqrt{1 - \frac{C_s^2}{æ(B)}} < 1.$$

5. Сходимость градиентного метода



6. **Теорема 3** (О сходимости метода градиентного спуска). *Пусть в релаксационном процессе направление спуска - направление антиградиента функции $\varphi(\boldsymbol{y})$ в точке $\boldsymbol{x}^k$:*

$$\boldsymbol{s}^k = -\boldsymbol{r}^k.$$

Тогда (4)*- метод градиентного спуска сходится (Temple 1939г., Канторович 1945г.)*

$$\|\boldsymbol{x}^k - \boldsymbol{x}\|_A \le \mu^k \|\boldsymbol{x}^0 - \boldsymbol{x}\|_A, \mu \le \sqrt{1 - \frac{1}{\text{æ}(B)}}.$$

7. Методы сопряжённых градиентов

Conjugate Gradient Methods

Опр. *Крыловское подпространство* $\mathcal{K}_k(B, \boldsymbol{y})$, *порождённое* $n \times n$*-матрицей* B *и вектором* $\boldsymbol{y} \in \mathbb{R}^n$ *– это линейная оболочка*

$$\mathcal{K}_k(B, \boldsymbol{y}) = \mathsf{span}\left\{\boldsymbol{y}, B\boldsymbol{y}, \ldots, B^{k-1}\boldsymbol{y}\right\}.$$

Ортогонализируем степенную последовательность $B^i\boldsymbol{y} \ldots$

8. **Теорема 4.** *Пусть $B^T = B$, векторы $\boldsymbol{g}^i = B^{i-1}\boldsymbol{y}$ линейно независимы для $1 \le i \le k$ и векторы $\boldsymbol{f}^1, ..., \boldsymbol{f}^k$ получены из векторов $\boldsymbol{g}^1, ..., \boldsymbol{g}^k$ с помощью процесса ортогонализации, тогда*

$$\boldsymbol{f}^1 = \boldsymbol{y}, \boldsymbol{f}^2 = B\boldsymbol{f}^1 - \alpha_1 \boldsymbol{f}^1,$$

$$\boldsymbol{f}^{i+1} = B\boldsymbol{f}^i - \alpha_i \boldsymbol{f}^i - \beta_{i-1}\boldsymbol{f}^{i-1},$$

где

$$\alpha_1 = \frac{\langle B\boldsymbol{f}^1, \boldsymbol{f}^1\rangle}{\langle \boldsymbol{f}^1, \boldsymbol{f}^1\rangle}, \quad \alpha_i = \frac{\langle B\boldsymbol{f}^i, \boldsymbol{f}^i\rangle}{\langle \boldsymbol{f}^i, \boldsymbol{f}^i)}, \quad i > 1,$$

$$\beta_{i-1} = \frac{\langle B\boldsymbol{f}^i, \boldsymbol{f}^{i-1}\rangle}{\langle \boldsymbol{f}^{i-1}, \boldsymbol{f}^{i-1}\rangle}.$$

9. Теорема верна и для энергетического скалярного произведения $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_B = \langle \boldsymbol{x}, B\, \boldsymbol{y} \rangle$; тогда $\{\boldsymbol{f}^k\}$ будут B-ортогональны, т.е.

$$\langle \boldsymbol{f}^i, \boldsymbol{f}^j \rangle_B = 0,\ i \neq j.$$

Матрица преобразования $\boldsymbol{y} \mapsto B\boldsymbol{y}$ в базисе $\boldsymbol{f}^1, \boldsymbol{f}^2, ..., \boldsymbol{f}^n$ пространства $\mathbb{R}^n$ трёхдиагональна:

$$B\boldsymbol{f}^i = \boldsymbol{f}^{i+1} + \alpha_i \boldsymbol{f}^i + \beta_{i-1} \boldsymbol{f}^{i-1}, \quad (i > 1)$$

если только $\boldsymbol{y}$ не принадлежит ни одному корневому подпространству матрицы B.

10. **Теорема 5** (Алгоритм метода сопряженных градиентов)**.** *Задан* $\boldsymbol{x}^0$; *вычисляем* $\boldsymbol{r}^0 = B\boldsymbol{x}^0 - \boldsymbol{b}$ *и полагаем* $\boldsymbol{s}^1 = \boldsymbol{r}^0$. *далее*

$$\begin{aligned} \boldsymbol{r}^n &= \boldsymbol{r}^{k-1} + \alpha_k B\boldsymbol{s}^k, \\ \boldsymbol{s}^{k+1} &= \boldsymbol{r}^k + \beta_k \boldsymbol{s}^k, \\ \boldsymbol{x}^k &= \boldsymbol{x}^{k-1} + \alpha_k \boldsymbol{s}^k, \end{aligned}$$

α_k *и* β_k *вычисляются по одной из формул*

$$\alpha_k = -\frac{(\boldsymbol{r}^{k-1}, \boldsymbol{r}^{k-1})}{(B\boldsymbol{s}^k, \boldsymbol{r}^{k-1})} = -\frac{(\boldsymbol{r}^{k-1}, \boldsymbol{s}^k)}{(B\boldsymbol{s}^k, \boldsymbol{s}^k)} < 0,$$

$$\beta_k = -\frac{(\boldsymbol{r}^k, B\boldsymbol{s}^k)}{(B\boldsymbol{s}^k, \boldsymbol{s}^k)} = \frac{(\boldsymbol{r}^k, \boldsymbol{r}^k)}{(\boldsymbol{r}^{k-1}, \boldsymbol{r}^{k-1})} > 0.$$

11. **Метод сопряжённых градиентов I**

$\boldsymbol{x}^0$ – задано; для всех $k \geq 0$ вычисляем

$$\boldsymbol{r}^k = \boldsymbol{\nabla}\varphi(\boldsymbol{y})\big|_{\boldsymbol{y}=\boldsymbol{x}^k} = B\boldsymbol{x}^k - \boldsymbol{b}; \quad \boldsymbol{s}^0 = \alpha_0 \boldsymbol{r}^0.$$

$\boldsymbol{s}^k = \alpha_k \boldsymbol{r}^{k-1} + \beta_k \boldsymbol{s}^{k-1}$. При $k > 0$:

$$(\alpha_k, \beta_k) = \underset{\alpha,\beta\in\mathbb{R}}{\operatorname{argmin}}\, \varphi(\boldsymbol{x}^k + \alpha\boldsymbol{r}^{k-1} + \beta\boldsymbol{s}^{k-1}),$$
$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \boldsymbol{s}^k;$$

Теорема 6. *Процесс релаксационен* $\varphi(\boldsymbol{x}^{k+1}) = \varphi(\boldsymbol{x}^k) + \frac{\alpha_k}{2}\|\boldsymbol{r}^k\|^2$, $\alpha_k \leq -\left(\frac{\|\boldsymbol{r}^k\|}{\|\boldsymbol{r}^k\|_B}\right)^2$, *и конечен* $\boldsymbol{x}^n = \boldsymbol{x}$, *причём*

$$\langle \boldsymbol{r}^i, \boldsymbol{r}^j \rangle = 0,\ \langle \boldsymbol{s}^i, \boldsymbol{s}^j \rangle_B = \langle B\boldsymbol{s}^i, \boldsymbol{s}^j \rangle = 0,\ \forall i \neq j.$$

12. Метод сопряжённых градиентов II

$\boldsymbol{x}^0$ – задано; для всех $k \geq 0$ вычисляем

$$\boldsymbol{s}^0 = \boldsymbol{r}^0, \text{ где } \boldsymbol{r}^k = \boldsymbol{\nabla}\varphi(\boldsymbol{y})\big|_{\boldsymbol{y}=\boldsymbol{x}^k};$$

$\beta_k = \operatorname*{argmin}_{0\leq\beta\in\mathbb{R}} \varphi(\boldsymbol{x}^k - \beta \boldsymbol{s}^{k-1})$; $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \beta_k \boldsymbol{s}^k$.

При $k > 0$, $\xi_k = -\frac{\langle \boldsymbol{r}^k, \boldsymbol{r}^k - \boldsymbol{r}^{k-1}\rangle}{\|\boldsymbol{r}^{k-1}\|^2}$, $\boldsymbol{s}^k = \boldsymbol{r}^k - \xi_k \boldsymbol{s}^{k-1}$.

Теорема 7. *Для сильно выпуклой* $\varphi(\boldsymbol{y})$:

$\varphi(\alpha\boldsymbol{y}+(1-\alpha)\boldsymbol{z})\leq\alpha\varphi(\boldsymbol{y})+(1-\alpha)\varphi(\boldsymbol{z})-\alpha(1-\alpha)\rho\|\boldsymbol{y}-\boldsymbol{z}\|^2$

градиент которой удовлетворяет условию

$$\|\boldsymbol{\nabla}\varphi(\boldsymbol{y}) - \boldsymbol{\nabla}\varphi(\boldsymbol{z})\| \leq L\|\boldsymbol{y} - \boldsymbol{z}\|,$$

процесс релаксационен и:

$\varphi(\boldsymbol{x}^k) - \varphi(\boldsymbol{x}) \leq [\varphi(\boldsymbol{x}^0) - \varphi(\boldsymbol{x})] \exp\{-\frac{k\rho^3}{2L(\rho+L)}\}$.

13. Методы симметризации системы

$$A\boldsymbol{x} \cong \boldsymbol{a}; \qquad \boldsymbol{a} \in \mathbb{R}^m, \boldsymbol{x} \in \mathbb{R}^n.$$

Метод AA^T-минимальных итераций:

$$\boldsymbol{x} = A^T\boldsymbol{u}, \boldsymbol{b} = \boldsymbol{a}, \varphi(\boldsymbol{y}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|^2 - \frac{1}{2}\|\boldsymbol{x}\|^2.$$

Метод A^TA-минимальных итераций:

$$\boldsymbol{u} = \boldsymbol{x}, \boldsymbol{b} = A^T\boldsymbol{a}, \varphi(\boldsymbol{y}) = \frac{1}{2}\|A\boldsymbol{y} - \boldsymbol{a}\|^2 - \frac{1}{2}\|\boldsymbol{a}\|^2.$$

14. Алгоритм CGLS [2]

Initialize

$$\boldsymbol{r}^{(0)} = \boldsymbol{b} - A\boldsymbol{x}^{(0)}, \boldsymbol{p}^{(0)} = \boldsymbol{s}^{(0)} = A^T\boldsymbol{r}^{(0)}, \gamma_0 = \|\boldsymbol{s}^{(0)}\|_m^2;$$

for $k = 0, 1, 2, \ldots$ **while** $\gamma_k > tol$ **compute**

$$\begin{aligned}
\boldsymbol{q}^{(k)} &= A\boldsymbol{p}^{(k)}, \\
\alpha_k &= \gamma_k / \|\boldsymbol{q}^{(k)}\|_m^2, \\
\boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{p}^{(k)}, \\
\boldsymbol{r}^{(k+1)} &= \boldsymbol{r}^{(k)} - \alpha_k \boldsymbol{p}^{(k)}, \\
\boldsymbol{s}^{(k+1)} &= A^T \boldsymbol{r}^{(k+1)}, \\
\gamma_{k+1} &= \|\boldsymbol{s}^{(k+1)}\|_n^2, \\
\beta_k &= \gamma_{k+1}/\gamma_k, \\
\boldsymbol{p}^{(k+1)} &= \boldsymbol{s}^{(k+1)} + \beta_k \boldsymbol{p}^{(k)}.
\end{aligned}$$

15. Точность алгоритма CGLS[3]

$$\frac{\|\boldsymbol{b} - A\boldsymbol{x}^{(k)} - \boldsymbol{r}^{(k)}\|_m}{\|A\|(\|\boldsymbol{x}\|_n + \|\boldsymbol{x}^{(0)}\|_n)} \leq$$
$$\leq \varepsilon_c \left((k+1+c) + k(10+2c) \max_{j \leq k} \frac{\|\boldsymbol{x}^{(j)}\|_n}{\|\boldsymbol{x}\|_n + \|\boldsymbol{x}^{(0)}\|_n} \right)$$

Здесь ε_c – машинная точность, c – связано с точностью вычисления произведения матрицы на вектор и обычно $c = m\sqrt{n}$.

16. Устойчивый алгоритм CGLS [2]

Initialize

$$\boldsymbol{p}^{(0)} = \boldsymbol{s}^{(0)} = A^T \left(\boldsymbol{b} - A\boldsymbol{x}^{(0)} \right), \gamma_0 = \|\boldsymbol{s}^{(0)}\|_m^2;$$

for $k = 0, 1, 2, \ldots$ **while** $\gamma_k > tol$ **real compute**

$$\begin{aligned}
\boldsymbol{q}^{(k)} &= A\boldsymbol{p}^{(k)}, \\
\alpha_k &= \gamma_k / \|\boldsymbol{q}^{(k)}\|_m^2, \\
\boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{p}^{(k)}, \\
\boldsymbol{s}^{(k+1)} &= \boldsymbol{s}^{(k)} - \alpha_k \left(A^T \boldsymbol{q}^{(k)} \right), \\
\gamma_{k+1} &= \|\boldsymbol{s}^{(k+1)}\|_n^2, \\
\beta_k &= \gamma_{k+1} / \gamma_k, \\
\boldsymbol{p}^{(k+1)} &= \boldsymbol{s}^{(k+1)} + \beta_k \boldsymbol{p}^{(k)}.
\end{aligned}$$

17. Предобуславливание в LSP

Preconditioned problem

$$\boldsymbol{x} = S^{-1}\boldsymbol{y} \Rightarrow AS^{-1}\boldsymbol{y} \cong \boldsymbol{b}; \boldsymbol{b} \in \mathbb{R}^m, \boldsymbol{y} \in \mathbb{R}^n. \qquad (6)$$

Условия стационарности в PCLSP

$$\boldsymbol{x} = S^{-1}\boldsymbol{y} \Rightarrow S^{-T}A^T(AS\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{0}.$$

CGLS с предобуславливанием:

Теорема 8. *[2]*

$$\left\|\boldsymbol{r} - \boldsymbol{r}^{(k)}\right\|_m < 2\left(\frac{æ(AS^{-1}) - 1}{æ(AS^{-1}) + 1}\right)^k \left\|\boldsymbol{r} - \boldsymbol{r}^{(0)}\right\|_m,$$

где $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$, $\boldsymbol{r}^{(k)} = \boldsymbol{b} - A\boldsymbol{x}^{(k)}$, $æ(AS^{-1})$ **- число обусловленности матрицы** AS^{-1}.

18. Алгоритм с предобуславливанием PCCGLS

Initialize

$$\boldsymbol{r}^{(0)} = \boldsymbol{b} - A\boldsymbol{x}^{(0)}, \boldsymbol{p}^{(0)} = \boldsymbol{s}^{(0)} = S^{-T}\left(A^T\boldsymbol{r}^{(0)}\right), \gamma_0 = \|\boldsymbol{s}^{(0)}\|_m^2;$$

for $k = 0, 1, 2, \ldots$ **while** $\gamma_k > tol$ **compute**

$$\begin{aligned}
\boldsymbol{t}^{(k)} &= S^{-1}\boldsymbol{p}^{(k)},\\
\boldsymbol{q}^{(k)} &= A\boldsymbol{t}^{(k)},\\
\alpha_k &= \gamma_k/\|\boldsymbol{q}^{(k)}\|_m^2,\\
\boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} + \alpha_k\boldsymbol{t}^{(k)},\\
\boldsymbol{r}^{(k+1)} &= \boldsymbol{r}^{(k)} - \alpha_k\boldsymbol{q}^{(k)},\\
\boldsymbol{s}^{(k+1)} &= S^{-T}\left(A^T\boldsymbol{r}^{(k+1)}\right),\\
\gamma_{k+1} &= \|\boldsymbol{s}^{(k+1)}\|_n^2,\\
\beta_k &= \gamma_{k+1}/\gamma_k,\\
\boldsymbol{p}^{(k+1)} &= \boldsymbol{s}^{(k+1)} + \beta_k\boldsymbol{p}^{(k)}.
\end{aligned}$$

19. Литература

1. **Лоусон Ч., Хенсон Р.** Численное решение задач метода наименьших квадратов. Пер. с англ. – М.: Наука. Гл. ред. физ.-мат. лит., 1986. — 232 с.

2. **Å. Björck.** Numerical Methods for Least Squares Problems. – SIAM, Philadelphia,1996.-427 p.

3. **Å. Björck, T. Elfving and Z. Strakos** Stability of conjugate gradient-type methods for linear least squares problems, Technical Report LiTH-MAT-R-1995-26, Department of Mathematics, Linköping University, (1994)

4. **Воеводин В.В., Кузнецов Ю.А.** Матрицы и вычисления. – М.: Наука, 1984. – 320с.

5. **Воеводин В.В.** Линейная алгебра. – М.: Наука, 1980. – 400с.

6. **Воеводин В.В.** Вычислительные основы линейной алгебры. – М.: Гл.ред.Физ.Мат,Лит., 1977. – 304с.

7. **Карманов В.Г.** Математическое программирование. – М.: ФизМатЛит, 2004. – 264с.