Лекция 16. Итерационные методы для решения линейных систем 3

Максим Рахуба

ФКН ВШЭ

Основы матричных вычислений 20/21

Май 24, 2021

План

Сходимость CG

GMRES

Предобуславливание

Итог

Метод сопряженных градиентов (CG)

$$\mathbf{A}\mathbf{x}_* = \mathbf{b}, \quad \mathbf{A} = \mathbf{A}^\intercal \succ 0.$$

Результаты из прошлой лекции

CG как оптимизация на подпространствах Крылова:

$$\mathbf{x}_k = \operatorname*{arg\,min}_{\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)} \|\mathbf{x} - \mathbf{x}_*\|_{\mathbf{A}}.$$

$= \{r_0, Ar_0, \dots, A^{k-1} r_0\}$

Отсюда мы получили "короткие" рекуррентные соотношения:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{x}_{k-1} + \alpha_k \mathbf{p}_k, \\ \mathbf{r}_k &= \mathbf{r}_{k-1} - \alpha_k \mathbf{A}\mathbf{p}_k, \\ \mathbf{p}_{k+1} &= \mathbf{r}_k + \beta_k \mathbf{p}_k. \end{aligned}$$

$r_k = b - Ax_k \iff$

$r_k \perp p_k \Rightarrow \alpha_k$

$p_k \perp_A p_{k+1} \Rightarrow \beta_k$

Сходимость CG

Лемма

Пусть $\mathbf{A} = \mathbf{A}^\top \succ 0$. *Тогда для CG справедлива следующая оценка:*

$$\underbrace{\|\mathbf{e}_k\|_{\mathbf{A}}}_{x_* - x_k} \le \min_{p_k : p_k(0)=1} \max_i |p_k(\overset{\text{с.з. } A}{\lambda_i})| \, \underbrace{\|\mathbf{e}_0\|_{\mathbf{A}}}_{x_* - x_0}.$$

$\square$ 1) $x_k = x_0 + y, \; y \in K_k(A, r_0) \Rightarrow x_k = x_0 + q_{k-1}(A) r_0 \Rightarrow$

$\Rightarrow r_k = b - A x_k = b - A(x_0 + q_{k-1}(A) r_0) = \underbrace{(b - A x_0)}_{r_0} - A q_{k-1}(A) r_0 =$

$= (I - A q_{k-1}(A)) r_0 = p_k(A) r_0, \quad p(0) = 1 \qquad r_0 = \sum_{i=1}^n c_i \overset{\text{с.в. } A}{v_i}$

$A e_k = A(x_* - x_k) = b - A x_k = r_k$

2) $\|e_k\|_A^2 = (e_k, A e_k) = (A^{-1} r_k, r_k) = (A^{-1} p_k(A) r_0, p_k(A) r_0)$

$= \left(\sum_{i=1}^n c_i A^{-1} p_k(A) v_i, \sum_{i=1}^n c_i p_k(A) v_i\right) = \left(\sum c_i \frac{p_k(\lambda_i)}{\lambda_i} v_i, \sum c_i p_k(\lambda_i) v_i\right)$

$= \sum_i \sum_j c_i c_j \frac{p_k(\lambda_i) p_k(\lambda_j)}{\lambda_i} (v_i, v_j) = \sum_{i=1}^n c_i^2 \frac{p_k(\lambda_i)^2}{\lambda_i} \le \max_i |p_k(\lambda_i)|^2 \underbrace{\sum_{i=1}^n \frac{c_i^2}{\lambda_i}}$

$(A^{-1} r_0, r_0) = \|e_0\|_A^2$

$r_0 = A e_0$

3) Берём min

[redacted]

Сходимость CG

$$\|e_k\|_A \leq \min_{p_k : p_k(0)=1} \max_i |p_k(\lambda_i)| \, \|e_0\|_A$$

$$\max_i |p_k(\lambda_i)| \leq \max_{\lambda \in [\lambda_n, \lambda_1]} |p_k(\lambda)|.$$

Сходимость CG

$$\|e_k\|_A \leq \min_{p_k : p_k(0)=1} \max_i |p_k(\lambda_i)| \, \|e_0\|_A$$

$$\max_i |p_k(\lambda_i)| \leq \max_{\lambda \in [\lambda_n, \lambda_1]} |p_k(\lambda)|.$$

Поэтому для CG справедлива оценка, аналогичная оценке для метода Чебышёва:

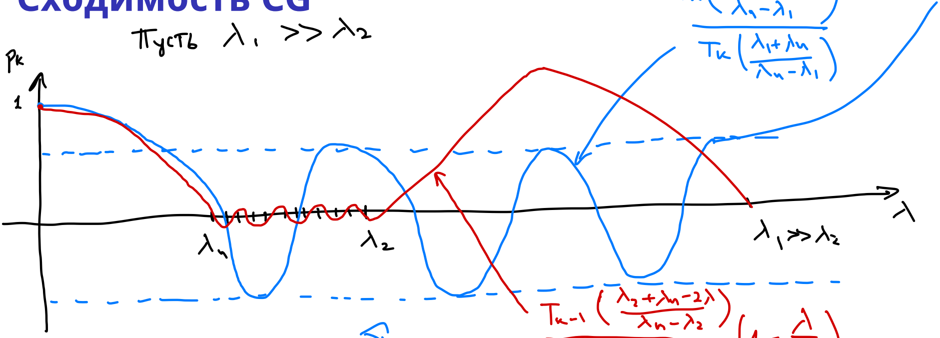
$\sqrt{\mathrm{cond}_2(A)}$ где $A = A^T > 0$

$$\|\mathbf{e}_k\|_{\mathbf{A}} \leq 2 \left(\frac{\underbrace{\sqrt{\frac{\lambda_1}{\lambda_n}}}_{=\sqrt{\mathrm{cond}_2(A)}} - 1}{\sqrt{\frac{\lambda_1}{\lambda_n}} + 1} \right)^k \|\mathbf{e}_0\|_{\mathbf{A}}.$$

Но можно лучше!

Сходимость CG

Пусть $\lambda_1 \gg \lambda_2$



$$\frac{T_k\left(\frac{\lambda_1+\lambda_n-2\lambda}{\lambda_n-\lambda_1}\right)}{T_k\left(\frac{\lambda_1+\lambda_n}{\lambda_n-\lambda_1}\right)}$$

$$\frac{T_{k-1}\left(\frac{\lambda_2+\lambda_n-2\lambda}{\lambda_n-\lambda_2}\right)}{T_{k-1}\left(\frac{\lambda_2+\lambda_n}{\lambda_n-\lambda_2}\right)}\left(1-\frac{\lambda}{\lambda_1}\right)$$

$\ll \sqrt{\lambda_1/\lambda_n}$

$\Downarrow$ т.к. $\left|1-\frac{\lambda}{\lambda_1}\right| \leq 1 \quad \forall \lambda \in [0, \lambda_1]$

$$\|e_k\|_A \leq 2\left(\frac{\sqrt{\frac{\lambda_2}{\lambda_n}}-1}{\sqrt{\frac{\lambda_2}{\lambda_n}}+1}\right)^{k-1}\|e_0\|_A$$

Т.е. CG "игнорирует" отдельные большие/маленькие с.з. Их может быть > 1

План

- Сходимость CG
- **GMRES**
- Предобуславливание
- Итог

Метод обобщенных минимальных невязок (GMRES)

симм. полож. опр.

$Ax_* = b$

$A^T A x_* = A^T b$

но $\mathrm{cond}_2(A^T A) = \mathrm{cond}_2(A)^2 >> \mathrm{cond}_2(A)$

$(Ax - b, Ax - b) = (A(x - x_*), A(x - x_*)) =$

$= ((x - x_*), A^T A(x - x_*)) = \|x - x_*\|_{A^T A}$

$\|$

Общий случай: $\mathbf{A} : \det(\mathbf{A}) \neq 0$.

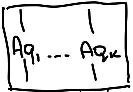
$$\mathbf{x}_k = \arg\min_{\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2.$$

“Короткие” формулы аналогичные CG получатся только при $\mathbf{A} = \mathbf{A}^\top$. В этом случае метод называется `minres`.

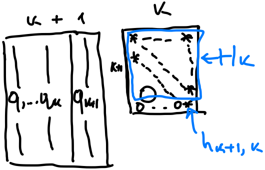
Ортогонализация подпространств Крылова

Нам понадобится $\mathbf{Q}_k = [\mathbf{q}_1, \ldots, \mathbf{q}_k]$ – ортогональный базис в $\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$, полученный с помощью ГШ ортогонализации системы векторов $\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{k-1}\mathbf{r}_0\}$.

Утверждение (соотношение Арнольди)



$$= \mathbf{A}\mathbf{Q}_k = \mathbf{Q}_{k+1}\widehat{\mathbf{H}}_k =$$



где $\widehat{\mathbf{H}}_k \in \mathbb{R}^{(k+1)\times k}$ *– верхняя хессенбергова.*

$$AQ_k = Q_k H_k + h_{k+1,k}\, q_{k+1} e_k^T$$

$$e_k^T = (0 \ldots 0\ 1)$$

Ортогонализация подпространств Крылова

Нам понадобится $\mathbf{Q}_k = [\mathbf{q}_1, \ldots, \mathbf{q}_k]$ – ортогональный базис в $\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$, полученный с помощью ГШ ортогонализации системы векторов $\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{k-1}\mathbf{r}_0\}$.

Утверждение (соотношение Арнольди)

$$\mathbf{A}\mathbf{Q}_k = \mathbf{Q}_{k+1}\widehat{\mathbf{H}}_k$$

где $\widehat{\mathbf{H}}_k \in \mathbb{R}^{(k+1)\times k}$ *– верхняя хессенбергова.*

$\square$ $\operatorname{Im}(AQ_k) \subset \operatorname{Im}(Q_{k+1})$, т.к. $A\,\mathrm{span}\{r_0, Ar_0, \ldots, A^{k-1}r_0\} = $

$= \mathrm{span}\{Ar_0, A^2 r_0, \ldots, A^k r_0\}$

Но $Aq_i \in \operatorname{Im}(Q_k)$, $i < k \Rightarrow Aq_k \in \operatorname{Im}(Q_{k+1})$

$$q_{k+1} = (I - Q_k Q_k^T) A q_k = A q_k - Q_k Q_k^T A q_k =$$

$$= A q_k - (q_1^T A q_k) q_1 - \ldots - (q_k^T A q_k) q_k$$

$$q_{k+1} = q_{k+1} / \|q_{k+1}\|_2$$

$$A q_k = \underbrace{\mathrm{const}}_{h_{k+1,k}} \cdot q_{k+1} + \overbrace{(q_1^T A q_k)}^{h_{1k}} q_1 + \ldots + \overbrace{(q_k^T A q_k)}^{h_{kk}} q_k$$

Ортогонализация подпространств Крылова

Следствие

Матрица $(\mathbf{Q}_k^\intercal \mathbf{A} \mathbf{Q}_k) \in \mathbb{R}^{k \times k}$ *является верхней хессенберговой, а в случае* $\mathbf{A} = \mathbf{A}^\intercal$ *— трехдиагональной.*

□

$$A Q_k = Q_k H_k + h_{k+1,k}\, q_{k+1} e_k^T$$

$$Q_k^T A Q_k = H_k + h_{k+1,k}\, \cancelto{0}{Q_k^T q_{k+1}}\, e_k^T$$

симм. верхнехесс. является Треугольной

■

GMRES

$$\mathbf{x}_k = \underset{\mathbf{x}\in\mathbf{x}_0+\mathcal{K}_k(\mathbf{A},\mathbf{r}_0)}{\arg\min} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2.$$

$$\|Ax-b\|_2 = \|A(x_0+Q_k c)-b\|_2 = \|AQ_k c - \underbrace{(b-Ax_0)}_{r_0}\|_2$$

$$= \|AQ_k c - r_0\|_2 = \|Q_{k+1}\hat{H}_k c - r_0\|_2 =$$

$$q_1 = \frac{r_0}{\|r_0\|_2} \Rightarrow r_0 = \|r_0\|_2 q_1 \qquad \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= \|Q_{k+1}\hat{H}_k c - \|r_0\|_2 q_1\|_2 = \|\cancel{Q_{k+1}}(\hat{H}_k c - \|r_0\|_2 e_1)\|_2$$

$$= \|\hat{H}_k c - \|r_0\|_2 e_1\|_2 \text{ — МНК}$$

$\hat{H}_k = UR$ — QR разлож. с помощью вращений Гивенса

получим $O(k^2)$ вместо $O(k^3)$

GMRES

Параметр `restart`

В отличии от cg и minres, в gmres на k-й итерации необходимо хранить $\mathcal{O}(k)$ векторов. Поэтому метод перезапускают каждые `restart` итераций, начиная с $\mathbf{x}_0 := \mathbf{x}_{\texttt{restart}}$.

План

Сходимость CG

GMRES

Предобуславливание

Итог

Предобуславливание

a.k.a. переобуславливание, preconditioning

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{b} \\ \mathbf{P}_L^{-1}\mathbf{A}\mathbf{x} &= \mathbf{P}_L^{-1}\mathbf{b} \\ (\mathbf{P}_L^{-1}\mathbf{A}\mathbf{P}_R^{-1})(\mathbf{P}_R\mathbf{x}) &= \mathbf{P}_L^{-1}\mathbf{b} \end{aligned}$$

Получаем эквивалентную предобусловленную систему:

$$(\mathbf{P}_L^{-1}\mathbf{A}\mathbf{P}_R^{-1})\mathbf{y} = \mathbf{P}_L^{-1}\mathbf{b}, \quad \mathbf{x} = \mathbf{P}_R^{-1}\mathbf{y},$$

где $\mathbf{P}_L, \mathbf{P}_R$ (левый и правый предобуславливатель):

- итерационный процесс с $\mathbf{P}_L^{-1}\mathbf{A}\mathbf{P}_R^{-1}$ сходится быстро: $\mathrm{cond}_2(\mathbf{P}_L^{-1}\mathbf{A}\mathbf{P}_R^{-1}) \approx 1$, с.з. были кластеризованы и т.д.
- матвеки $\mathbf{P}_L^{-1}\mathbf{x}$ и $\mathbf{P}_R^{-1}\mathbf{x}$ быстро вычислять.

Примеры предобуславливателей

1. Пусть $\mathbf{A} = \mathbf{L} + \mathbf{D} + \mathbf{U}$.
   - Якоби ($\mathbf{P} = \mathbf{D}$)
   - Гаусс-Зейдель ($\mathbf{P} = \mathbf{L} + \mathbf{D}$)
   - послед. верхней релаксации ($\mathbf{P} = \omega^{-1}(\mathbf{D} + \omega\mathbf{L})$), ...
2. Блочные версии предобуславливателей из 1. Например, блочный Якоби:

$$\mathbf{P} = \begin{bmatrix} \mathbf{A}_{11} & & \\ & \ddots & \\ & & \mathbf{A}_{bb} \end{bmatrix}, \quad \text{где} \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \dots & \mathbf{A}_{1b} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{b1} & \dots & \mathbf{A}_{bb} \end{bmatrix}$$

3. Неполное разложение Холецкого или LU (incomplete LU, ILU). Отбрасываем заполнения:
   - По шаблону (e.g., оставляем ненул. элем. как в $\mathbf{A}^k$, $k \geq 1$).
   - и/или пренебрегая элементами $\leq$ заданной величины (`droptol` параметр).
   - ...

Предобуславливание CG

Вообще говоря, нельзя использовать только левый или только правый предобуславливатели для CG:

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

так как $(\mathbf{P}^{-1}\mathbf{A})^{\mathsf{T}} \neq \mathbf{P}^{-1}\mathbf{A}$.

Предобуславливание CG

Вообще говоря, нельзя использовать только левый или только правый предобуславливатели для CG:

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

так как $(\mathbf{P}^{-1}\mathbf{A})^\intercal \neq \mathbf{P}^{-1}\mathbf{A}$.

Формально запишем $\mathbf{P} = \mathbf{C}^\intercal\mathbf{C}$ и рассмотрим:

$$\mathbf{C}^{-\intercal}\mathbf{A}\mathbf{C}^{-1}\mathbf{y} = \mathbf{C}^{-1}\mathbf{b}.$$

Итоговые формулы CG можно привести к виду только с $\mathbf{P}^{-1}$ (ДЗ).

План

Сходимость CG

GMRES

Предобуславливание

Итог

Итог

$A = A^T$

- нет: $A^T x$ доступно?
  - нет: память проблема?
    - нет: gmres с большим restart.
    - да: gmres с мал. знач. restart или bicgstab
  - да: QMR (квази мин. невязок)
- да: $A > 0$ (или $A < 0$)
  - нет: minres
  - да: CG (Если известно λ_1, λ_n и не хочется считать скалярные произведения, то можно попробовать метод Чебышёва)

Литература

- Saad Y., Iterative methods for sparse linear systems. SIAM, 2003.
- Лебедев В.И. Функциональный анализ и вычислительная математика, Физматлит, 2005.
- Тыртышников Е. Е. Методы численного анализа. – Академия, 2007. – 320 с.
- Demmel J. Applied numerical linear algebra. SIAM, 1997