SPEEDING UP KRYLOV SUBSPACE METHODS FOR COMPUTING f(A)b VIA RANDOMIZATION

Команда: speedruners

Участники: Владимир Добрыгин, Газиз Абдрахман, Владислав Гаухов

https://github.com/loadi1/nla_ploj2

Основная идея

- В статье предлагается использовать рандомизированные методы подпространства. Подпространство Крылова K_m(A,b) обычно строится с использованием алгоритма Арнольди, но этот процесс становится вычислительно дорогостоящим по мере увеличения размера подпространства (числа итераций, m). Чтобы преодолеть это, авторы предлагают использовать неортонормированные базисы для подпространства Крылова и использовать рандомизированные методы для ускорения вычислений.

Как считать по Арнольди

$$AU_m = U_m K_m + k_{m+1,m} u_{m+1} e_m^T = U_{m+1} \underline{K}_m, \qquad \underline{K}_m = \begin{bmatrix} K_m \\ k_{m+1,m} e_m^T \end{bmatrix} \in \mathbb{R}^{(m+1)\times m}$$

где U_m - ортонормированный базис, K_m - это сжатие A в m-е подпространство Крылова, а u_{m+1} - единичный вектор, который расширяет U_m до U_{m+1} .

Как считают в статье

$$AV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^T = V_{m+1} \underline{H}_m, \qquad \underline{H}_m = \begin{bmatrix} H_m \\ h_{m+1,m} e_m^T \end{bmatrix} \in \mathbb{R}^{(m+1)\times m},$$

где V_m - НЕортонормированный базис, H_m - верхняя матрица Гессенберга, а u_{m+1} - единичный вектор, который расширяет U_m до U_{m+1} .

Концепция наброска

- Ключевая идея заключается в использовании матрицы эскиза $\Theta \in s \times n$, где s выбирается таким образом, что $m < s \ll n$.
- Эта матрица эскиза используется для проецирования как текущего базиса, так и нового вектора в пространство меньшей размерности. В этом сокращенном пространстве процесс ортогонализации намного дешевле. Алгоритм работает следующим образом:

1. На каждом шаге i генерируется новый вектор w_i путем умножения матрицы A последним вектором в текущем базисе V_{i-1}.

2. Сохраняется эскизная версия базиса, $S_{i-1} = \Theta V_{i-1}$.

3. Вычисляется эскиз нового вектора $p_i := \Theta\, w_i$.

4. Ортогонализация выполняется для малой матрицы $[\, S_{i-1} \& \, p_i \,]$ с использованием одного шага Грама-Шмидта, в результате чего получается обновленный схематичный базис S_i с ортонормированными столбцами.

5. Коэффициенты, полученные на этом этапе ортогонализации, сохраняются в верхней треугольной матрице R.

6. Эти коэффициенты используются для приблизительной ортогонализации w_i относительно предыдущих векторов в фактическом (не зарисованном) пространстве для получения v_i.

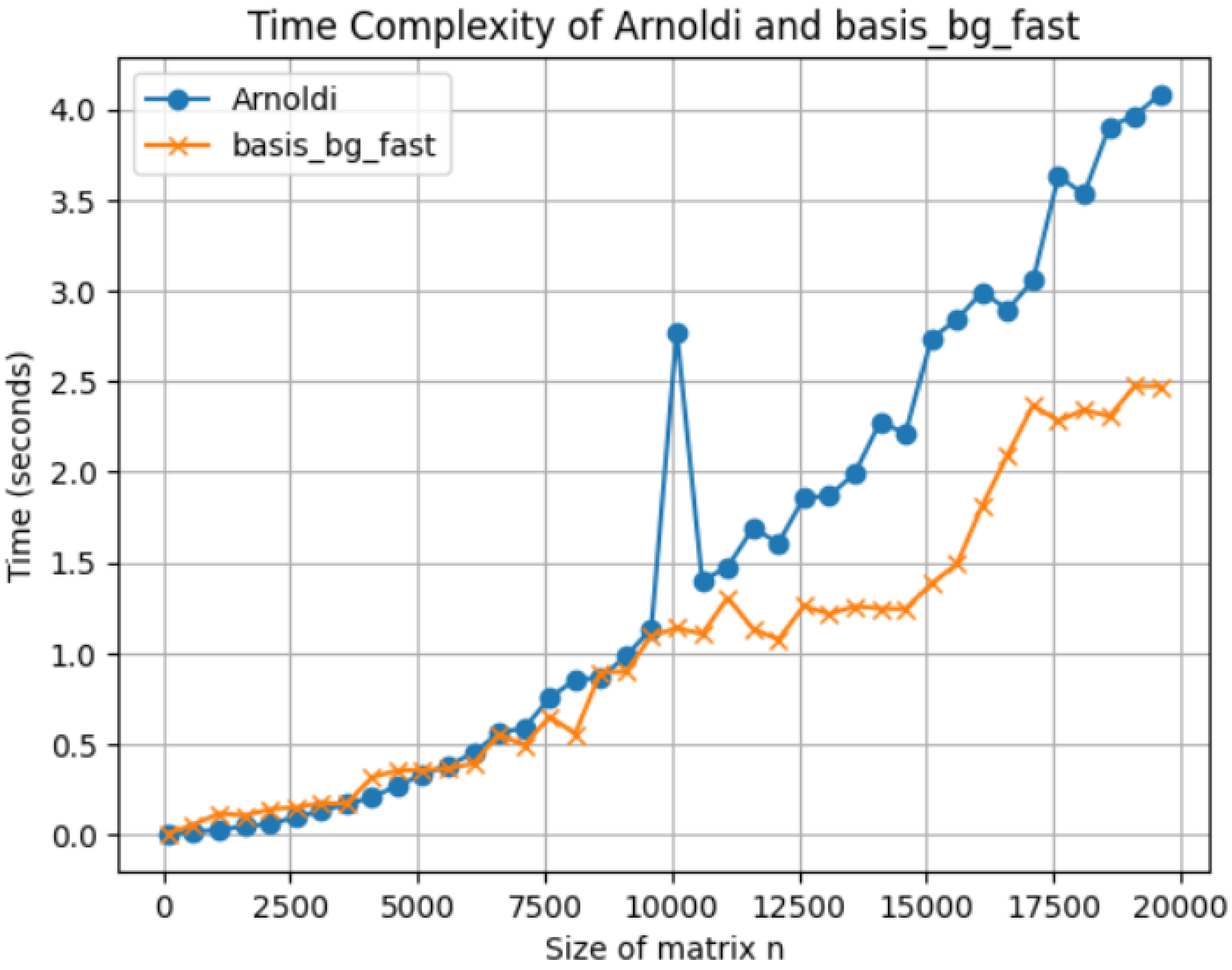
7. Новый вектор v_i масштабируется так, что его эскиз имеет единичную длину.

- Этот процесс гарантирует, что, хотя фактические базисные векторы не являются точно ортогональными, их эскизные версии являются таковыми. Если матрица эскиза выбрана соответствующим образом, можно показать, что результирующий базис V_{m+1} хорошо обусловлен.
- Авторы предлагают использовать рандомизированное преобразование Адамара с подвыборкой для Θ, которое позволяет быстро вычислять эскизы за $\mathcal{O}$(nlog n) времени.
- Экономия вычислений этого метода по сравнению с алгоритмом Арнольди обусловлена тем фактом, что на каждой итерации требуется только одно умножение с текущим базисом V_{i-1} (по сравнению с двумя в алгоритме Арнольди). Это может привести к значительному ускорению, особенно когда матрица A разрежена.

Метод усеченной ортогонализации является примером компромисса между эффективностью вычислений и численной точностью. Ограничивая процесс ортогонализации, он позволяет ускорить вычисления, что может быть особенно полезно при решении очень больших задач, где полная ортогонализация невозможна.

- Однако в тексте также подчеркивается ключевой недостаток этого подхода: база может стать плохо подготовленной или даже с численным ранжированием недостаточной всего после нескольких итераций. Для решения этой проблемы в тексте предлагается процесс "отбеливания", который включает в себя мониторинг номера условия базиса и выполнение приблизительной ортогонализации, когда это необходимо.
- В процессе отбеливания используется матрица эскизов для сохранения сжатого представления основы. При обнаружении плохого кондиционирования вычисляется QR-факторизация нарисованного базиса, и исходный базис обновляется до значения $V_i\ R_i^{-1}$, которое должно быть хорошо кондиционировано. Матрица H_{i-1} в соотношении, подобном Арнольди, обновляется соответствующим образом.

	Форма V_m	Сложность составления наброска
Арнольди	nm^2	-
Алгоритм из статьи	nm	$nm \log m$



Time Complexity of Arnoldi and basis_bg_fast
Arnoldi
basis_bg_fast
Time (seconds)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0
2500
5000
7500
10000
12500
15000
17500
20000
Size of matrix n