

Abschlussprojekt Numerische Mathematik:
Cholesky-Zerlegung von Skyline-Matrizen

Ida Hönigmann

Fabian Dopf

3. Februar 2022

Inhaltsverzeichnis

Aufgabe 1:	Aufwandsordnung numerischer Verfahren	2
Teilaufgabe 1a	2
Teilaufgabe 1b	2
Teilaufgabe 1c	3
Aufgabe 2:	Cholesky-Verfahren und Skyline-Matrizen	4
Teilaufgabe 2a	4
Teilaufgabe 2b	6
Aufgabe 3:	Pseudocode für Cholesky-Zerlegung von Skyline-Matrizen	8
Teilaufgabe 3a	8
Teilaufgabe 3b	9
Aufgabe 4:	Aufwand des Algorithmus und Verhalten in Spezialfällen	10
Teilaufgabe 4a	10
Teilaufgabe 4b	11
Aufgabe 5:	Implementierung des Algorithmus und empirische Aufwandsschätzung	13
Appendix A:	Python Implementierung	14

Aufgabe 1: Aufwandsordnung numerischer Verfahren

Wir betrachten ein abstraktes numerisches Verfahren, das für $N \in \mathbb{N}$ Eingabedaten eine Laufzeit von $y_N \in \mathbb{R}_+$ hat. Man sagt, das Verfahren habe Aufwandsordnung $p > 0$, falls eine Konstante $C > 0$ existiert, sodass $y_N \leq CN^p$ für alle $N \in \mathbb{N}$.

Teilaufgabe 1a

Die Aufwandsordnung lässt sich über die Folge $\{p_N\}_{N \in \mathbb{N}}$ mit

$$p_N = \frac{\log(y_{2N}) - \log(y_N)}{\log(2)} \text{ für } N \in \mathbb{N} \quad (1)$$

quantifizieren. Beachten Sie, dass die Bestimmung von p_N die Verfügbarkeit von zwei aufeinanderfolgenden Folgengliedern y_N und y_{2N} erfordert. Verwenden Sie den Ansatz $y_N = CN^p$ und leiten Sie die Formel in 1 her!

Beweis. Sei $(y_N)_{N \in \mathbb{N}}$ die Laufzeit eines numerischen Verfahrens. Wir verwenden den Ansatz $y_N = CN^p$ für ein $C > 0$ und $N \in \mathbb{N}$.

Durch Umformen ergibt sich

$$\begin{aligned} \log(y_{2N}) - \log(y_N) &= \log\left(\frac{y_{2N}}{y_N}\right) = \log\left(\frac{C(2N)^p}{C \cdot N^p}\right) = \log(2^p) = p \log(2), \text{ d.h.} \\ p &= \frac{\log(y_{2N}) - \log(y_N)}{\log(2)}. \end{aligned}$$

Verhält sich also y_N asymptotisch wie CN^p , so konvergiert p_N gegen p .

□

Teilaufgabe 1b

Sei $\{\delta_N\}_{N \in \mathbb{N}} \subseteq \mathbb{R}$ eine Nullfolge, d.h. es gilt $\delta_N \rightarrow 0$ für $N \rightarrow \infty$. Weiters verhalte sich die Laufzeit wie $y_N = (C + \delta_N)N^p$ mit $C > 0$. Zeigen Sie, dass die Folge $\{p_N\}_{N \in \mathbb{N}}$ gegen p konvergiert, d.h. es gilt $p_N \rightarrow p$ für $N \rightarrow \infty$.

Beweis. Zuerst berechnen wir einen Grenzwert, den wir in späterer Folge verwenden werden.

$$\lim_{n \rightarrow \infty} \log\left(\frac{C + \delta_{2N}}{C + \delta_N}\right) \stackrel{(1)}{=} \log\left(\lim_{n \rightarrow \infty} \frac{C + \delta_{2N}}{C + \delta_N}\right) = \log\left(\frac{\lim_{n \rightarrow \infty} C + \delta_{2N}}{\lim_{n \rightarrow \infty} C + \delta_N}\right) \stackrel{(2)}{=} \log\left(\frac{C}{C}\right) = \log(1) = 0$$

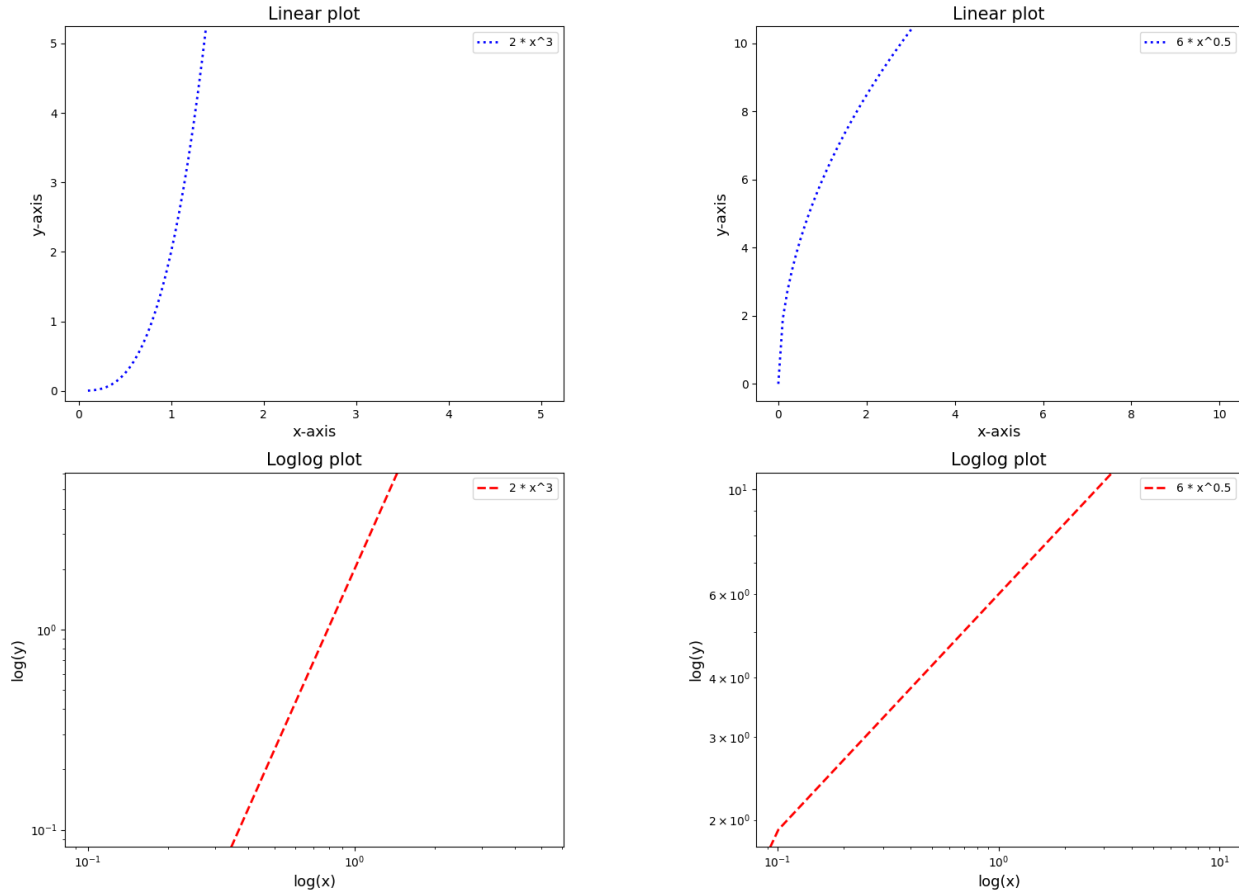
Hierbei folgt (1) aus der Stetigkeit des Logarithmus und (2) aus der Voraussetzung $\delta_N \rightarrow 0$ (womit auch $\delta_{2N} \rightarrow 0$).

Setzen wir nun die Voraussetzung $y_N = (C + \delta_N)N^p$ in Gleichung 1 ein, so erhalten wir

$$\begin{aligned} p_N &= \frac{\log(y_{2N}) - \log(y_N)}{\log(2)} = \frac{\log((C + \delta_{2N})(2N)^p) - \log((C + \delta_N)N^p)}{\log(2)} = \frac{\log\left(\frac{(C + \delta_{2N})(2N)^p}{(C + \delta_N)N^p}\right)}{\log(2)} \\ &= \frac{\log\left(\frac{(C + \delta_{2N})2^p}{(C + \delta_N)}\right)}{\log(2)} = \frac{p \log(2) + \log\left(\frac{C + \delta_{2N}}{C + \delta_N}\right)}{\log(2)} = p + \frac{\log\left(\frac{C + \delta_{2N}}{C + \delta_N}\right)}{\log(2)} \xrightarrow{n \rightarrow \infty} p + 0 = p \end{aligned}$$

Zusammenfassend gilt nun $\lim_{n \rightarrow \infty} p_n = p$, was zu zeigen war.

□



(a) Vergleich der beiden Graphen der Gleichung $y = 2x^3$. Im Loglog plot ist zu erkennen, dass die Steigung der Geraden 3 bei gleichen Achsenskalierungen ist, also gilt $p = 3$.

(b) Vergleich der beiden Graphen der Gleichung $y = 6x^{0.5}$. Im Loglog plot ist zu erkennen, dass bei der Funktion bei $x = 1$ den Funktionswert 6 annimmt, also gilt $c = 6$.

Abbildung 1: Zwei unterschiedliche Funktionen der Form $y = cx^p$, jeweils in einem linear skalierten und einem Loglog Plot.

Teilaufgabe 1c

In sogenannter doppelt logarithmischer Darstellung (log-log Plots) wird für beide Koordinatenachsen eine logarithmische Skalierung verwendet, d.h. sowohl die waagrechte als auch die senkrechte Koordinatenachse wird logarithmisch unterteilt. Wie werden Potenzfunktionen der Form $y = cx^p$ in einem log-log Plot dargestellt? Wie können Sie die Ordnung p und die Konstante $c > 0$ aus einem log-log Plot von $y = cx^p$ direkt auslesen?

Ausarbeitung. Berechnen wir für eine Funktion der Form $y = cx^p$ die Werte $X = \log x$ und $Y = \log y$, so erhalten wir

$$Y = \log y = \log(cx^p) = \log c + p \log x = \log c + pX \quad (2)$$

was der Gleichung einer Geraden entspricht.

Um den Wert von c abzulesen, kann man den Funktionswert bei 1 ablesen, da $f(1) = c \cdot 1^p = c$. Um p abzulesen, kann man die Steigung der Geraden ansehen, wenn beide Achsen "gleich" skaliert, da p die Steigung der Geraden in Gleichung 2 ist.

Aufgabe 2: Cholesky-Verfahren und Skyline-Matrizen

Eine Matrix $A \in \mathbb{R}^{n \times n}$ heißt Skyline-Matrix, falls es für $l = 1, \dots, n$ Zahlen $p_l, q_l \in \mathbb{N}_0$ gibt, sodass für die i -te Zeile und j -te Spalte von A gilt:

- $A_{i,k} = 0$ für $k < i - p_i$,
- $A_{k,j} = 0$ für $k < j - q_j$,

Folgendes Beispiel illustriert diese Aussage:

$$A = \begin{pmatrix} 1 & & & & \\ & 1 & & 2 & 2 \\ & & 1 & 3 & 3 \\ 1 & 2 & 3 & 14 & 18 \\ & 4 & 5 & 29 & 48 \end{pmatrix}.$$

Teilaufgabe 2a

Beweisen Sie, dass das Cholesky-Verfahren genau dann wohldefiniert ist (d.h. es wird nicht durch Null dividiert oder die Wurzel aus einer negativen Zahl gezogen), wenn die Matrix $A \in \mathbb{R}^{n \times n}$ symmetrisch und positiv definit ist.

Beweis der Existenz der Cholesky-Zerlegung. Wir wiederholen zuerst die relevanten Definitionen.

- Eine Matrix $A \in \mathbb{R}^{n \times n}$ heißt symmetrisch, falls $A = A^T$.
- Eine Matrix $A \in \mathbb{R}^{n \times n}$ heißt positiv definit, falls $\forall u \in \mathbb{R}^n \setminus \{0\} : u^T A u > 0$.

Wir zeigen, dass für jede SPD Matrix $A \in \mathbb{R}^{n \times n}$ eine untere Dreiecksmatrix $L \in \mathbb{R}^{n \times n}$ existiert, sodass $LL^T = A$, durch vollständige Induktion nach n :

- **Induktionsanfang:** $n = 1$

Wenn $A := (a_{11}) \in \mathbb{R}^{1 \times 1}$ eine beliebige SPD Matrix ist, so folgt für $u := (1) \in \mathbb{R}^1 \setminus \{0\}$

$$0 < u^T A u = (1) (a_{11}) (1) = a_{11}.$$

Da also $a_{11} > 0$ gilt, ist $L := (\sqrt{a_{11}}) \in \mathbb{R}^1$ wohldefiniert. Dann gilt

$$LL^T = (\sqrt{a_{11}}) (\sqrt{a_{11}}) = (a_{11}) = A.$$

- **Induktionsvoraussetzung:** $\forall A \in \mathbb{R}^{n-1 \times n-1}$ SPD $\exists L \in \mathbb{R}^{n-1 \times n-1}$ untere Dreiecksmatrix : $LL^T = A$.
- **Induktionsschritt:** $n - 1 \implies n$

Sei $A \in \mathbb{R}^{n \times n}$ eine symmetrische und positiv definite Matrix. Wir definieren eine Matrix $B \in \mathbb{R}^{(n-1) \times (n-1)}$, einen Vektor $a \in \mathbb{R}^{n-1}$ und eine Zahl $\alpha \in \mathbb{R}$ durch

$$\forall i, j \in \{1, \dots, n-1\} : B_{i,j} := A_{i,j}, \quad \forall i \in \{1, \dots, n-1\} : a_i := A_{i,n}, \quad \alpha := A_{n,n}.$$

Zusammengefasst gilt nun wegen der Symmetrie von A

$$A = \begin{pmatrix} B & a \\ a^T & \alpha \end{pmatrix}$$

B ist eine symmetrische, positiv definite Matrix aus $\mathbb{R}^{(n-1) \times (n-1)}$, daher existiert laut Induktionsvoraussetzung eine untere Dreiecksmatrix $P \in \mathbb{R}^{(n-1) \times (n-1)}$ mit $PP^T = B$.

Da B positiv definit ist und somit regulär ist, folgt die eindeutige Existenz eines Vektors $l \in \mathbb{R}^{n-1}$ der die Gleichung $Pl = a$ erfüllt.

Wir wollen nun $\beta \in \mathbb{R}$ so definieren, dass $\beta = \sqrt{\alpha - l^T l}$. Dazu müssen wir sicherstellen, dass $\alpha - l^T l > 0$.

Wenn wir die Definition von l verwenden und Umformen erhalten wir

$$\begin{aligned}\alpha - l^T l &= \alpha - (P^{-1}a)^T (P^{-1}a) = \alpha - a^T (P^{-1})^T P^{-1}a \\ &= \alpha - a^T (PP^T)^{-1}a = \alpha - a^T B^{-1}a.\end{aligned}$$

Da A positiv definit ist ergibt sich

$$0 < \begin{pmatrix} -B^{-1}a \\ 1 \end{pmatrix}^T \underbrace{\begin{pmatrix} B & a \\ a^T & \alpha \end{pmatrix}}_{=A} \begin{pmatrix} -B^{-1}a \\ 1 \end{pmatrix} = \alpha - a^T B^{-1}a.$$

Also ist $\beta := \sqrt{\alpha - l^T l}$ wohldefiniert. Umgeformt gilt nun $l^T l + \beta^2 = \alpha$.

Definieren wir nun $L \in \mathbb{R}^{n \times n}$ durch

$$L = \begin{pmatrix} P & 0 \\ l^T & \beta \end{pmatrix},$$

Dann gilt

$$LL^T = \begin{pmatrix} P & 0 \\ l^T & \beta \end{pmatrix} \begin{pmatrix} P^T & l \\ 0 & \beta \end{pmatrix} = \begin{pmatrix} PP^T & Pl \\ l^T P^T & l^T l + \beta^2 \end{pmatrix} = \begin{pmatrix} PP^T & Pl \\ (Pl)^T & l^T l + \beta^2 \end{pmatrix} = \begin{pmatrix} B & a \\ a^T & \alpha \end{pmatrix} = A,$$

was zu zeigen war.

□

Beweis der Wohldefiniertheit. Angenommen das Cholesky-Verfahren ist wohldefiniert. Sei $A \in \mathbb{R}^{n \times n}$ beliebig und $L \in \mathbb{R}^{n \times n}$ die untere Dreiecksmatrix der Cholesky-Zerlegung.

Aus der Eigenschaft $A = LL^T$ der Cholesky-Zerlegung ergibt sich

$$A^T = (LL^T)^T = L^{TT} L^T = LL^T = A \quad (3)$$

woraus sich die Symmetrie von A ergibt.

Sei $x \in \mathbb{R}^n \setminus \{0\}$ beliebig. Wir definieren $u := L^T x \in \mathbb{R}^n$. Da L eine reguläre Matrix ist gilt $u \neq 0$ und

$$x^T A x = x^T L L^T x = u^T u > 0. \quad (4)$$

Insgesamt zeigen Gleichungen 3 und 4, dass A eine SPD Matrix ist.

Wenn umgekehrt A eine SPD Matrix ist, so zeigt der Beweis der Existenz der Cholesky-Zerlegung, dass die Cholesky-Zerlegung existiert und an keiner Stelle durch Null dividiert oder die Wurzel einer negativen Zahl gezogen wird. □

Aus der Vorlesung ist folgender Algorithmus zur Berechnung einer Cholesky-Zerlegung bekannt:

Definition 1 (Algorithmus zur Berechnung einer Cholesky-Zerlegung).

Input: SPD Matrix $A \in \mathbb{R}^{n \times n}$

- (1) for $k = 1:n$
- (2) $L_{kk} = \sqrt{A_{kk} - \sum_{j=1}^{k-1} (L_{kj})^2}$
- (3) for $i = k+1:n$
- (4) $L_{ik} = (A_{ik} - \sum_{j=1}^{k-1} L_{ij} L_{kj}) / (L_{kk})$
- (5) end loop
- (6) end loop

Output: untere Dreiecksmatrix $L \in \mathbb{R}^{n \times n}$

Beweis der Wohldefiniertheit des Algorithmus aus Definition 1. Aus der Vorlesung ist bereits bekannt, dass auf Werte L_{ik} nur zugreift, falls diese bereits berechnet wurden.

Um die Wohldefiniertheit zu zeigen, muss einerseits wegen Zeile (2) und andererseits wegen Zeile (4) gelten, dass $A_{kk} - \sum_{j=1}^{k-1} (L_{kj})^2 > 0$ für alle $k \in \{1, \dots, n\}$.

Aus der Eigenschaft $A = LL^T$ ergibt sich

$$A_{kk} = \sum_{j=1}^k (L_{kj})^2 = (L_{kk})^2 + \sum_{j=1}^{k-1} (L_{kj})^2 \geq \sum_{j=1}^{k-1} (L_{kj})^2$$

und daraus $A_{kk} - \sum_{j=1}^{k-1} (L_{kj})^2 \geq 0$.

Angenommen $A_{kk} - \sum_{j=1}^{k-1} (L_{kj})^2 = 0$ und somit auch $L_{kk} = 0$. Dann wäre L nicht regulär, da die Determinante gleich Null ist. Gemeinsam mit $A = LL^T$ ist das jedoch ein Widerspruch zu A ist positiv definit (also auch regulär). Das heißt es muss $A_{kk} - \sum_{j=1}^{k-1} (L_{kj})^2 > 0$ gelten, was zu zeigen war. \square

Teilaufgabe 2b

Beweisen Sie, dass die Besetzungsstruktur der Cholesky-Zerlegung der Skyline-Matrix A erhalten bleibt, d.h. dass auch die untere Dreiecksmatrix L eine geeignete Bandstruktur aufweist.

Beweis. Aus der Vorlesung ist bekannt, dass der Algorithmus aus Definition 1 die Choleksy-Zerlegung einer symmetrischen, positiv definiten Matrix berechnet.

Weiters ist bekannt, dass im Algorithmus auf Werte L_{ik} nur zugegriffen wird, falls diese bereits berechnet wurden.

Sei $A \in \mathbb{R}^{n \times n}$ eine beliebige symmetrische, positiv definite Skyline Matrix mit $p_l = q_l \in \{1, \dots, l\}$, d.h.

$$A_{ik} = 0 \text{ für } i, k \in \{1, \dots, n\} \text{ mit } k < i - p_i.$$

Sei $L \in \mathbb{R}^{n \times n}$ die untere Dreiecksmatrix aus der Cholesky-Zerlegung. Sei $i \in \{1, \dots, n\}$ beliebig.

Wir zeigen

$$\forall k \in \{1, \dots, i - p_i - 1\} : L_{ik} = 0 \quad (5)$$

durch vollständige Induktion nach k :

Falls $\{1, \dots, i - p_i - 1\} = \emptyset$ ist nichts zu zeigen.

1. **Induktionsanfang:** $k = 1$

Für $k = 1$ gilt

$$1 \leq i - p_i - 1 \iff 1 < i - p_i \implies A_{i1} = 0.$$

L_{i1} wird in Zeile 4 des Algorithmus wie folgt berechnet und danach nicht mehr geändert:

$$L_{i1} = \frac{A_{i1} - \sum_{j=1}^0 L_{ij} L_{1j}}{L_{11}} = \frac{0}{L_{11}} = 0$$

2. **Induktionsvoraussetzung:** Es gelte $\forall j \in \{1, \dots, k-1\} : L_{ij} = 0$.

3. **Induktionsschritt:** $k-1 \implies k$

Für k gilt

$$k \leq i - p_i - 1 \iff k < i - p_i \implies A_{ik} = 0.$$

L_{ik} wird in Zeile 4 des Algorithmus wie folgt berechnet und danach nicht mehr geändert:

$$L_{ik} = \frac{A_{ik} - \sum_{j=1}^{k-1} L_{ij} L_{kj}}{L_{kk}} = \frac{0 - \sum_{j=1}^{k-1} 0 \cdot L_{kj}}{L_{kk}} = 0$$

Dadurch ist gezeigt, dass die Skyline-Form in der unteren Dreieckshälfte der Cholesky-Zerlegung erhalten bleibt. \square

Aufgabe 3: Pseudocode für Cholesky-Zerlegung von Skyline-Matrizen

Verwenden Sie den Cholesky-Algorithmus aus der Vorlesung. Entwerfen Sie jeweils einen Pseudocode, der für eine Skyline-Matrix:

Teilaufgabe 3a

möglichst effizient die Struktur erkennt.

Ausarbeitung.

```
1  to_skyline
2  Input: n times n matrix A
3      values = empty list
4
5      for i from 0 to n:
6          up_branch = i-th column of A from diagonal to top
7          remove all entries in up_branch outside of skyline-branch
8
9          left_branch = i-th row from diagonal to left
10         remove all entries in left_branch outside of skyline-branch
11
12         append up_branch and left_branch to values
13
14  Output: values
```

Listing 1: Strukturerkennung einer nicht notwendigerweise symmetrischen Skyline-Matrix

```
1  to_spd_skyline
2  Input: n times n SPD matrix A
3      values = empty list
4
5      for i from 0 to n:
6          branch = i-th column of A from diagonal to top
7          remove all entries in up_branch outside of skyline-branch
8
9          append branch to values
10
11  Output: values
```

Listing 2: Strukturerkennung einer symmetrischen positiv definiten Skyline-Matrix

Teilaufgabe 3b

die Cholesky-Zerlegung berechnet.

Ausarbeitung.

```
1  cholesky
2  Input: n times n SPD matrix A
3  L = zero matrix of dimension n by n
4
5  for k from 0 to n:
6      s = 0
7      for j from 0 to k:
8          s += L[k, j] ** 2
9
10     L[k, k] = sqrt(A[k, k] - s)
11
12     for i from k+1 to n:
13         s = 0
14         for j from 0 to k:
15             s += L[i, j] * L[k, j]
16
17         L[i, k] = (A[i, k] - s) / L[k, k]
18
19 Output: matrix L
```

Listing 3: Algorithmus für die Cholesky Zerlegung einer Matrix aus der Vorlesung

```
1  cholesky_skyline
2  Input: values as returned by to_spd_skyline
3  L = zero matrix of dimension n by n
4
5  max_width = maximum branch length
6
7  for k from 0 to n:
8      start_idx = index of first entry not equal to zero
9
10     # realize sum as dot product of two vectors
11     v = vector of k-th row from start_idx to k of matrix L
12     s = v * v
13     e = element in position k, k from original matrix
14     L[k, k] = sqrt(e - s)
15
16     for i from k + 1 to min(k + max_width, n):
17         if value inside of the branch length
18             # realize sum as dot product of two vectors
19             v = vector of i-th row from start_idx to k of matrix L
20             w = vector of k-th row from start_idx to k of matrix L
21             s = v * w
22             e = element in position i, k from original matrix
23             L[i, k] = (e - s) / L[k, k]
24
25 Output: matrix L
```

Listing 4: Optimierter Algorithmus für die Cholesky Zerlegung einer Skyline-Matrix

Aufgabe 4: Aufwand des Algorithmus und Verhalten in Spezialfällen

Teilaufgabe 4a

Sei $A \in \mathbb{R}^{n \times n}$ eine Skyline-Matrix. Welchen Aufwand haben Ihre Algorithmen aus Aufgabe 3 in Abhängigkeit von der Größe n der Eingabedaten und Skyline-Indices $p_l = q_l$?

Ausarbeitung.

- `to_skyline` hat wegen der Schleife in Zeile 5 und den beiden Zeilen 7 und 10 Aufwand

$$\sum_{l=1}^n (n - p_l) + (n - q_l) = 2n^2 - \sum_{l=1}^n p_l + q_l.$$

- best case: Im besten Fall ist $p_l = q_l = l$ also maximal, mit Aufwand

$$2n^2 - \sum_{l=1}^n 2l = 2n^2 - 2 \frac{n(n+1)}{2} = 2n^2 - (n^2 + n) = n^2 - n$$

- worst case: Im schlechtesten Fall ist $p_l = q_l = 0$ also minimal, mit Aufwand

$$2n^2 - \sum_{l=1}^n 0 = 2n^2$$

- `to_spd_skyline` hat wegen der Schleife in Zeile 5 und Zeile 7 Aufwand

$$\sum_{l=1}^n (n - p_l) = n^2 - \sum_{l=1}^n p_l$$

- best case: Im besten Fall ist $p_l = l$ also maximal, mit Aufwand

$$n^2 - \sum_{l=1}^n l = n^2 - \frac{n(n+1)}{2} = n^2 - \frac{1}{2}n^2 - \frac{1}{2}n = \frac{1}{2}(n^2 - n)$$

- worst case: Im schlechtesten Fall ist $p_l = 1$ also minimal, mit Aufwand

$$n^2 - \sum_{l=1}^n 1 = n^2 - n$$

- `cholesky` hat wegen den Schleifen in Zeile 5, 7, 12 und 14 Aufwand

$$\begin{aligned} \sum_{k=1}^n (k + \sum_{i=k+1}^n k) &= \sum_{k=1}^n (k + k(n-k)) = \sum_{k=1}^n k + \sum_{k=1}^n kn - \sum_{k=1}^n k^2 \\ &= \frac{n(n+1)}{2} + \frac{n^2(n+1)}{2} - \frac{n(n+1)(2n+1)}{6} = \frac{1}{6}n^3 + \frac{1}{2}n^2 + \frac{1}{3}n \end{aligned}$$

also $\mathcal{O}(n^3)$.

- `cholesky_skyline` Zeilen 12 und 21 haben Aufwand $\mathcal{O}(p_k)$, da die beiden Vektoren Länge

$$k - \text{start_idx} + 1 = k - (k - p_k + 1) + 1 = p_k$$

haben und sind somit, neben Zeile 5 mit Aufwand von $\mathcal{O}(n)$, die aufwändigsten Zeilen in der Funktion.

Durch das if-Statement in Zeile 17 wird sichergestellt, dass der Aufwand nur anfällt, falls nach Aufgabe 2b das Ergebnis nicht trivial ist.

Die Einschränkung in Zeile 16 auf den Bereich `k+1 to k+max_width` statt `k+1 to n` ermöglicht es viele der if-Statements zu ersparen, die es dem Python-Interpreter erschweren die Ausführung zu optimieren¹.

Insgesamt ist der Aufwand also

$$n + \sum_{k=1}^n \sum_{l=1}^{p_k} p_k = n + \sum_{k=1}^n p_k^2$$

was

- im besten Fall $n + \sum_{k=1}^n 1 = 2n = \mathcal{O}(n)$,
- im Fall, dass $p_l \approx \sqrt{n}$ ungefähr $n + \sum_{k=1}^n (\sqrt{n})^2 = n + n^2 = \mathcal{O}(n^2)$ und
- im schlechtesten Fall $n + \sum_{k=1}^n k^2 = n + \frac{n(n+1)(2n+1)}{6} = \frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{7}{6}n = \mathcal{O}(n^3)$ ist.

Teilaufgabe 4b

Betrachten Sie Matrizen mit den Besetzungsstrukturen

$$\begin{pmatrix} * & * & * & * & * \\ * & * & & & \\ * & & * & & \\ * & & & * & \\ * & & & & * \end{pmatrix} \quad \text{bzw.} \quad \begin{pmatrix} * & & & & * \\ & * & & & * \\ & & * & & * \\ & & & * & * \\ * & * & * & * & * \end{pmatrix}$$

Welche Besetzungsstruktur hat die Cholesky-Zerlegung für beide Matrizen? Was könnte man machen, um für Matrizen mit der "linken" Besetzungsstruktur die Cholesky-Zerlegung effizienter zu berechnen?

Ausarbeitung.

- Die linke Matrix ist vollbesetzt als Skyline-Matrix, in dem Sinne, dass

$$\forall k \in \{1, \dots, n\} : p_k = q_k = k$$

also immer den maximal möglichen Wert annimmt.

- Die rechte Matrix hat die Skyline-Indizes

$$\forall k \in \{1, \dots, n-1\} : p_k = q_k = 1 \quad \text{und} \quad p_n = q_n = n$$

und ist daher nach Aufgabe 4a effizienter in der Berechnung der Cholesky-Zerlegung.

Eine effiziente Berechnung der Cholesky-Zerlegung der linken Matrix erhält man mit folgender Überlegung: Definieren wir eine Abbildung die einer Matrix die entlang der Gegendiagonale gespiegelte Matrix zuordnet:

$$\sigma : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n} \quad (\sigma(A))_{i,j} := A_{n+1-j, n+1-i}$$

σ hat folgende Eigenschaften: Für beliebige $A, B \in \mathbb{R}^{n \times n}$ gilt

- Die Umkehrabbildung $\sigma^{-1} = \sigma$, da

$$\sigma(\sigma(A))_{ij} = \sigma(A)_{n+1-j, n+1-i} = A_{n+1-(n+1-i), n+1-(n+1-j)} = A_{i,j}$$

¹If-Statements ermöglichen es nicht die Instruction Pipeline vollständig auszunützen und verlangsamen dadurch die Ausführung. In modernen Systemen wird daher sogenannte Branch Prediction durchgeführt die versucht vorherzusagen ob die Bedingung wahr ist.

- Verträglich mit Transponieren

$$\sigma(A^T)_{i,j} = A_{n+1-j,n+1-i}^T = A_{n+1-i,n+1-j} = \sigma(A)_{j,i} = \sigma(A)_{i,j}^T$$

- Verträglich mit Multiplizieren

$$\begin{aligned} (\sigma(A)\sigma(B))_{i,j} &= \sum_{k=1}^n \sigma(A)_{i,k} \sigma(B)_{k,j} = \sum_{k=1}^n A_{n+1-k,n+1-i} B_{n+1-j,n+1-k} \\ &= \sum_{k=1}^n B_{n+1-j,k} A_{k,n+1-i} = (BA)_{n+1-j,n+1-i} = \sigma(BA)_{i,j} \end{aligned}$$

da $k \mapsto n+1-k$ eine Permutation von $\{1, \dots, n\}$ ist.

- Dreiecksmatrix bleibt erhalten

Wenn A eine untere Dreiecksmatrix ist, also $A_{i,j} = 0$ für $i < j$, so folgt, dass $\sigma(A)$ auch eine untere Dreiecksmatrix ist, da

$$i < j \iff n+1-j < n+1-i.$$

- Symmetrie bleibt erhalten

Wenn A symmetrisch ist, so ist auch $\sigma(A)$ symmetrisch, da

$$\sigma(A)_{i,j} = A_{n+1-j,n+1-i} = A_{n+1-i,n+1-j} = \sigma(A)_{j,i}.$$

Sei A eine Matrix von der linken, ungünstigen Art. Dann ist $\sigma(A)$ von der rechten Art, und ist es ist daher möglich die Cholesky-Zerlegung von $\sigma(A)$ effizient zu berechnen. Sei L eben diese Zerlegung von $\sigma(A)$, d.h. $LL^T = \sigma(A)$.

Nach dem oben gezeigten gilt nun

$$\sigma(L)^T \sigma(L) = \sigma(L^T) \sigma(L) = \sigma(LL^T) = \sigma(\sigma(A)) = A,$$

womit wir eine Cholesky-Zerlegung von A erhalten.

Aufgabe 5: Implementierung des Algorithmus und empirische Aufwandsschätzung

Implementieren Sie Ihren modifizierten Cholesky-Algorithmus in Python und weisen Sie empirisch nach, dass der Aufwand linear in n wächst. Vergleichen Sie die Performance Ihrer Implementierung mit der Python-Funktion `scipy.linalg.cholesky`, wobei die Skyline-Matrix A als vollbesetzte Matrix gespeichert ist.

Ausarbeitung. Die gesamte Python Implementierung kann auf https://github.com/idahoenigmann/numerical_analysis_final_project gefunden werden. Die Implementierung von `to_spd_skyline` und `cholesky_skyline` ist außerdem im Anhang zu finden.

Der Aufwand von `cholesky_skyline` hängt, wie in Aufgabe 3a begründet, stark von den Werten p_l ab. Grafik 2 zeigt die Zeitdauer der Cholesky Zerlegung für unterschiedliche durchschnittliche p_l Werte.

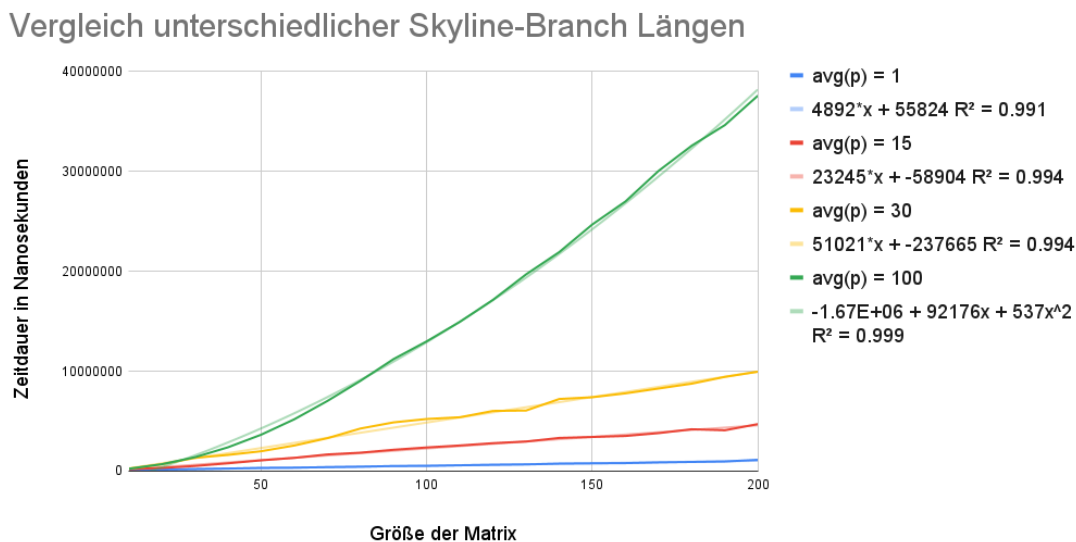


Abbildung 2: Für durchschnittliche p_l Werte bis 30 verhält sich der Aufwand fast linear. Bei großen durchschnittlichen p_l Werten, wie etwa 100 ist der Aufwand bereits polynomial.

Folgende Grafik vergleicht die Laufzeiten der unterschiedlichen Implementierungen aus Aufgabe 3b.

Da große Teile von `numpy` in C geschrieben sind, ist es nicht verwunderlich, dass `numpy.linalg.cholesky` deutlich schneller läuft als `cholesky_skyline`.

Vergleich unterschiedlicher Implementierungen

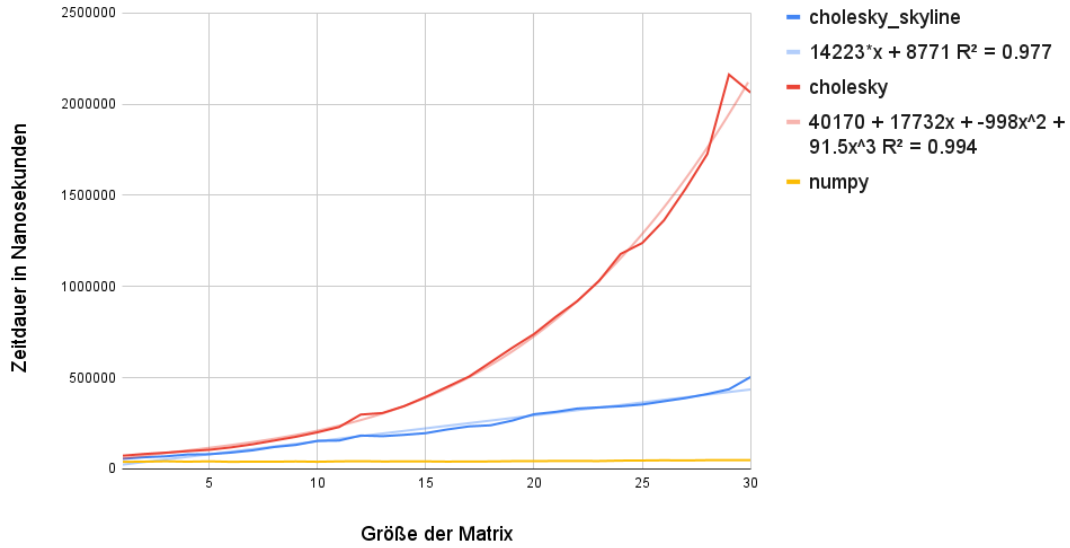


Abbildung 3: Vergleich des Aufwand der unterschiedlichen Implementierungen für Matrizen bis Größe 30×30 und durchschnittlichem p_l Wert von 10. Der Aufwand von `cholesky_skyline` ist deutlich besser, als jener von `cholesky`, allerdings ist `numpy.linalg.cholesky`, mit fast konstantem Aufwand, die schnellste Variante.

Appendix A: Python Implementierung

```

1  import numpy as np
2  import matrix_utils
3
4
5  class SPDSkylineMatrix:
6      values = list()
7
8      def __init__(self, matrix):
9          if not matrix_utils.is_symmetrical(matrix):
10             raise Exception("matrix must be symmetric")
11          if not matrix_utils.is_positive_definite(matrix):
12             raise Exception("matrix must be positive definite")
13
14          for i in range(len(matrix)):
15              branch = matrix[:, i][:i + 1].T.tolist()
16              branch.reverse()
17              while len(branch) > 0 and branch[-1] == 0:
18                  branch.pop(-1)
19
20              self.values.append(branch)
21
22      def __getitem__(self, item):
23          row_idx, col_idx = item
24          row_idx, col_idx = min(row_idx, col_idx), max(row_idx, col_idx)
25
26          if row_idx < 0 or col_idx > len(self.values):
27              raise Exception("index out of bounds")
28
29          col = self.values[col_idx]
30
31          if col_idx - row_idx < len(col):
32              return col[col_idx - row_idx]

```

```

33     else:
34         return 0
35
36     def to_matrix(self):
37         matrix = np.zeros((len(self.values), len(self.values)))
38
39         for i in range(len(self.values)):
40             for j in range(len(self.values[i])):
41                 matrix[i - j, i] = self.values[i][j]
42                 matrix[i, i - j] = self.values[i][j]
43
44         return matrix
45
46     def cholesky(self):
47         l = np.zeros((len(self.values), len(self.values)))
48
49         max_width = np.max(list(len(lst) for lst in self.values))
50
51         for k in range(len(self.values)):
52             start_idx = k - len(self.values[k]) + 1
53             s = np.dot(l[k][start_idx:k], l[k][start_idx:k])
54             l[k, k] = np.sqrt(self[k, k] - s)
55
56             for i in range(k + 1, min(k + max_width, len(self.values))):
57                 if k > i - len(self.values[i]):
58                     s = np.dot(l[i][start_idx:k], l[k][start_idx:k])
59                     l[i, k] = (self[i, k] - s) / l[k, k]
60
61         return l

```

Listing 5: Auszug der Implementierung der Umwandlung in Skyline-Form sowie der Cholesky Zerlegung für die Skyline-Form