Convex Optimization and Applications - Project

1. **מגישים:**

- **שם: גל ברזנסקי ת״ז: 205485816**
- **שם: אנטון לובמן ת״ז: 321123788**
- **שם: דור נריה ת״ז: 205439920**

2. **תיאור הבעיה:**

ביצענו פירוק צ׳ולסקי לבעיה המקורית וקיבלנו את הבעיה הבאה:

$$\min_{K\succ 0 \;\; K \text{ is } k-Band} \quad Tr(SK) - \log(|K|)$$

$$\Leftrightarrow \min_{\substack{L \text{ is } k-Band \\ L \text{ is lower triangular}}} \quad Tr(SLL^T) - \log(|LL^T|)$$

$$\Leftrightarrow \min_{\substack{L \text{ is } k-Band \\ L \text{ is lower triangular}}} \quad Tr(SLL^T) - \log(|L|^2)$$

$$\Leftrightarrow \min_{\substack{L \text{ is } k-Band \\ L \text{ is lower triangular}}} \quad Tr(SLL^T) - \sum_{i}^{n} \log\left(\left(L_{i,i}\right)^2\right)$$

כאשר L היא מטריצה משולשית תחתונה ו $k - Band$.

הסיבה שיכולנו לבצע את הפירוק הנ״ל היא כיון שנתון ש K היא PSD ולכן לא יכול להיות שKה שמביאה למינימום היא לא הפיכה שכן במקרה זה הדטרמיננטה תהיה 0 ו $logdet$ ישאף לאינסוף.

הוכחה שפירוק צ׳ולסקי יוצא $\boldsymbol{k - band}$: ראה נספח.

3. **האלגוריתם:**

בחרנו להשתמש באלגוריתם $gradient\ descent$.

כאשר בכל איטרציה אנו מטילים למרחב ה $k - Band$ ע״י איפוס האינדקסים הרלוונטיים בגראדיינט שקיבלנו.

בנוסף נשים לב שאנחנו לא צריכים לדאוג לכך המטריצה שלנו היא PSD מכיוון שהמטריצה שנחזיר היא LL^T שהיא כמובן PSD.

4. **התאמת האלגוריתם לשיפור ביצועים:**

התאמנו את האלגוריתם ע״י כך שחישבנו את הגראדיינט בהתחשב בתכונות של L. וקיבלנו שע״פ הגרסה השקולה האחרונה לבעייה הגראדיינט לפי L הוא:

$$\frac{\partial f}{\partial L} = 2 \cdot S \cdot L - diag\left(\frac{2}{L_{i,i}}\right) = 2(S \cdot L - diag(L)^{-1})$$

כאשר $diag(L)$ היא מטריצה אלכסונית עם האלכסון הראשי של L. נשים לב שכדי לחשב את הגראדיינט הנ״ל אין צורך להפוך מטריצה.
בנוסף גודל הצעד מחושב באמצעות Line Search ע״פ האלגוריתם שמופיע בספר הקורס.

<u>חילוק לתת בעיות:</u>

חילקנו את המטריצה לבלוקים בגודל $k+1$ על האלכסון, ועשינו אופטימיזציה עם GD עבור הבלוק המתאים ב S. ראה הסבר בנספח על אופן הפירוק לתתי בעיות.
לאחר מכן חיברנו את כל המטריצות למטריצה גדולה ודרסנו את הערכים החופפים.
ניחוש ראשוני- אם הבלוק של S באותו צעד הפיך אז אנו לוקחים את:

$$P\left(cholesky(S^{-1})\right)$$

כאשר P היא ההטלה ל- $k - Band$ המתאים.
אם הבלוק איננו הפיך אז אנחנו לוקחים את:

$$P\left(cholesky\left(\frac{1}{S}\right)\right)$$

כאשר $\left[\frac{1}{S}\right]_{i,j} = \frac{1}{[S]_{i,j}}$.

5. התחשבות במימד של S וזמן ריצה כללי:

האלגוריתם אינו מושפע כמעט מהמימד של S אלא בעיקר מהפרמטר k של $k - Band$.

בפירוט: מספר תתי הבעיות אותו נפתור מושפע מ-S שכן יהיו $n - k + 1$ תתי בעיות,

אולם כל תת בעיה בפני עצמה תלויה רק ב k. סה״כ זמן ריצה לכל תת בעיה הוא $O(k^3)$. ראה נספח.

סה״כ $O(nk^3)$, זאת בניגוד לפיתרון ללא חלוקה לתתי בעיות שהיה לוקח $O(n^3)$. ולכן עבור k קטן החלוקה לתתי בעיות שיפרה משמעותית את זמן הריצה.

נספח

1. **הוכחה שפירוק צ׳ולסקי יוצא $\boldsymbol{k-band}$:** נראה שבהינתן מטריצה L משולשית תחתונה ו $K-band$. נקבל המטריצה $K=LL^T$ היא גם $k-Band$ עם אותו k. עבור בה״כ $i>j+k$ מתקיים:

$$[LL^T]_{i,j} = \sum_{m=1}^{n} L_{i,m}\cdot L_{j,m}$$

כעת עבור $1\le m\le j$ $L_{i,m}=0,$ כיוון ש $m+k\le j+k<i$ ומכאן $i-m>k$.

ועבור $m>j$ מתקיים $L_{j,m}=0$ כיוון ש L משולשית תחתונה.

סה״כ קיבלנו

$$[LL^T]_{i,j} = 0 \quad \forall \quad |i-j|>k \quad \Rightarrow \quad K=LL^T\ is\ k-Band$$

2. **הסבר הפירוק לתתי בעיות**:

נחלק את S לבלוקים בגודל $k+1$ על האלכסון. ונכפיל ב L_1 בגודל המתאים:

$$Tr\left(S^1\cdot L_1\cdot {L_1}^T\right)-\sum_i^n \log\left(\left(L_{i,i}\right)^2\right)=\sum_i\sum_j\sum_l S^1_{ij}\cdot L_{lj}\cdot L_{i,l}-\sum_i^n \log\left(\left(L_{i,i}\right)^2\right)$$

$$Tr(S\cdot L\cdot L^T)-\sum_i^n \log\left(\left(L_{i,i}\right)^2\right)=\sum_i\sum_j\sum_l S_{ij}\cdot L_{lj}\cdot L_{i,l}-\sum_i^n \log\left(\left(L_{i,i}\right)^2\right)=$$

$$\rightsquigarrow \sum_p\left(\sum_i\sum_j\sum_l \left(S_p\right)_{ij}\cdot (L_p)_{lj}\cdot (L_p)_{i,l}-\sum_i^n \log\left(\left((L_p)_{i,i}\right)^2\right)=\right)$$

כאשר S_p,L_p הם תת מטריצות המהוות בלוקים בגודל $k+1$

3. **זמן ריצה לתת בעיה:**

בחישוב הניחוש הראשוני אם התת מטריצה S_p הפיכה אזי מבצעים פירוק צ׳ולסקי על ההופכית- $O(k^3)$

וכן בחישוב הגראדיינט מתבצע כפל מטריצות $S\cdot L$ שהוא גם מתבצע ב $O(k^3)$ ולכן סה״כ $O(k^3)$ לכל תת בעיה.