# Review 186: HiPPO: Recurrent Memory with Optimal Polynomial Projections

**Paper: https://arxiv.org/abs/2008.07669v2**

https://arxiv.org/abs/2008.07669

הגענו למאמר השני בסדרה - המאמר הזה חשוב מאוד כי הוא מפתח בסיס מתמטי מוצק המשמש כל המודלים מבוססים על מערכות דינמיות לינאריות כולל כמובן ממבה. המאמר הזה קצת (די הרבה) כבד מתמטית אך אנסה לעשות כמיטב יכולתי כדי להעביר לכם את המסר העיקרי שהוא מביא איתו.

בסקירה הקודמת דיברנו על איך ניתן לבנות וקטור זיכרון (m(t בעל יכולת לשחזר פונקצית קלט (u(x ל-x מאינטרוול ; כאן t מסמן גודל חלון הקשר (כלומר אורך הזיכרון). פונקצית (m(t ממודלת על ידי מערכת דינמית לינארית ושילובה עם פולינומי Legendre משחזר לנו את הקלט u. נעיר שאנו עובדים עם הגרסאות הדיסקרטיות של המודלים האלו שהן בעצם נוסחת נסיגה עבור סדרת וקטורי הזיכרון m\_t.

המאמר המסוקר מנסח מסגרת מתמטית כללית עבור בעיית ייצוג הזיכרון של פונקצית קלט (u(x בתחום . והנה מתחיל הסיבוך: קודם כל פולינומי Legendre הם מקרה פרטי של פונקציות אורתוגונליות במרחב הילברט (יותר נכון מרחב פונקציונלי L של לבג - המקרה הפרטי של הילברט) המצויד בנוסף בפונקציית מידה mu. אוקיי, מה הדבר הזה אומר בעצם? ממש בגדול זה מרחב של פונקציות שהמכפלה הפנימית ביניהן מוגדרת בתור אינטגרל של מכפלתן תחת מידה mu (במקרה הפשוט ביותר מידה mu שווה ל 1 זהותית ואנו מקבלים אינטגרל Riemann רגיל של המכפלה אבל עבור mu מורכבים יותר כמו Riemann-Stieltjes). פונקציות אורתוגונלית במרחב החמוד הזה מוגדרות בתור אלו שהמכפלה הפנימית שלהן שווה ל 0 (תחת מידה mu). פולינומי Legenge הן אורתוגונליים תחת מידה mu השווה ל- ב- ואפס בכל מקום אחר.

אז נניח שיש לנו N פונקציות אורתוגונליות במרחבנו החמוד. ועכשיו המטרה היא לתאר את הקלט (u(x ב- על ידי . כלומר אנו רוצים לבנות סכום ממושקל (u\*(x של עם מקדמים מסוימים (שימו לב שעבור t-ים שונים מקבלים וקטורי מקדמים שונים וכך שיש לנו כאן פונקציה וקטורית של המקדמים התלויה ב-t).

כלומר (u\*(x צריך לקרב בצורה טובה את הקלט (u(x (כלומר למזער שגיאה ביניהן ב-). והדיוק מחושב בתור אינטגרל של ההפרש הריבועי בין (u\*(x ו- (u(x תחת מידה mu (כאמור היא שווה ל- ב- עבור כל x ואפס בכל מקום אחר עבור פולינומי Legendre אבל כמובן קיימות עוד אפשרויות). איך נחשב מקדמים הממזערים את ההפרש הזה? לא כזה מסובך: מקדם i שווה למכפלה פנימית (=אינטגרל) בין פונקציה מספר i לפונקצית קלט u תחת אותה מידה mu.

עכשיו איך כל זה קשור למערכות דינמיות לינאריות החמודות שלנו? מתברר כי מערכת דינמית לינארית שתיארנו בסקירה הקודמת עבור וקטור (m(t מתארת את המקדמים של ייצוג הקלט באמצעות N פולינומי Legendre אורתוגונליים תחת מידה mu שהגדרנו לפני. ו-N זה המימד של וקטור הזיכרון (m(t תחת מידה mu הדורשת קרבה אחידה (=זכרון אחיד) בין u\* ו- u ב- .

אם נגדיר מידה mu להיות פונקציה (exp(x-t עבור t נתון, מערכת דינמית לינארית אחרת תתאר לנו מקדמים של פולינומי Laguerre (אורתוגונליים תחת mu הזו). שימו לב שמידה זו מגדירה זיכרון הדועך מעריכית כלומר ככל שעובר הזמן מזמן הנוכחי t, הזיכרון הולך ונהיה מעומעם יותר.

בנוסף המאמר מדבר גם על שיטות דיסקרטיזציה של מערכת דינמית זו וגם דן בקשר בינה לבין RNNs.

אוקיי, עכשיו סיכום במשפט אחד של המאמר הדי כבד הזה. המחברים בנו מסגרת מתמטית למידול בעיית הזיכרון של פונקציית קלט שישמש אותנו מאחורי הקלעים לבניית מודלי attention כל הדרך לממבה.