קורס : קודים לתיקון שגיאות ושימושיהם במדעי המחשב
מרצה : ד״ר קלים יפרמנקו
סמסטר : סתיו תשפ״א
תאריך : 29/11/2020

הרצאה 13

Justesen Codes

מבוא

Justesen Codes הינה קבוצה של קודים לתיקון שגיאות בעלי $R, \delta, |\Sigma|$ קבועים. לפני שקודים אלה הומצאו, לא היו ידועים קודים שעבורם כל שלושת הפרמטרים הנ״ל קבועים. כתוצאה מהגילוי שלהם, הומצאו קודים נוספים בעלי תכונה זו, למשל Expander Codes.

Justesen Codes בעלי שימושים חשובים במדעי המחשב, לדוגמה בבנייה של מרחבי מדגם בעלי Small Bias.

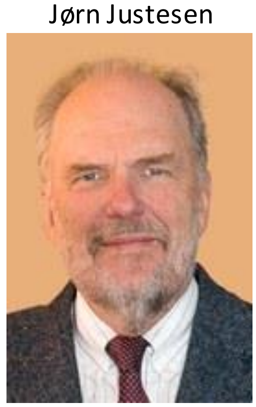
Justesen Codes הינם שרשור של קודי Reed Solomon עם קודי Wozencraft Ensemble, כאשר הראשונים בעלי R, δ קבועים על חשבון $|\Sigma|$ ליניארי באורך ההודעה, והאחרונים בעלי $R, |\Sigma|$ קבועים, אך δ קבוע רק ברוב הקודים במשפחה זו.

בניגוד לקודי Concatenation רגילים בהם הקוד הפנימי זהה עבור כל סימבול של מילת קוד מהקוד החיצוני, Justesen Codes מקודדים בעזרת RS, ואז מקודדים כל סימבול בעזרת קוד Wozencraft Ensemble שונה.

<u>הערה</u>: Justesen Code הינו קוד ״מפורש חזק״.

<u>הגדרה</u>: קוד $C: F^k \rightarrow F^n$ נקרא ״מפורש חזק״ אם לכל $1 \leq i \leq n, 1 \leq j \leq k$ קיים אלגוריתם הרץ בזמן $poly(logn)$ המחשב את $M[i,j]$, כאשר M הינה המטריצה היוצרת של C.

Jørn Justesen



הגדרה

Justesen Code הינו שרשור של $C_{out} = [n_{out}, k_{out}, d_{out}]_{2^m}$ עם $C_{in}^i = [n_{in}, k_{in}, d_i]_2$ שונים עבור $1 \leq i \leq n_{out}$. קוד זה, נסמנו $C^* = C_{out} \diamond \{C_{in}^1, \dots, C_{in}^{n_{out}}\}$, מקבל הודעה בינארית $m = m_1, \dots, m_{k_{out}}$ ומחשב תחילה את $C_{out}(m)$. לאחר מכן, השרשור מתאים לכל סימבול קוד מהקבוצה C_{in}^i.

$$\boxed{m_1, \dots, m_{k_{out}} \rightarrow C_{out}(m_1, \dots, m_{k_{out}}) = c_1, c_2, \dots, c_{n_{out}} \rightarrow C_{in}^1(c_1), C_{in}^2(c_2), \dots, C_{in}^{n_{out}}(c_{n_{out}})}$$

<u>הערה</u>: C_{out} הינו קוד RS ו $\{C_{in}^1, \dots, C_{in}^{n_{out}}\}$ הינה קבוצה של קודי Wozencraft Ensemble שונים.

<u>משפט:</u> אם $\{C_{in}^1, \dots, C_{in}^{n_{out}}\}$ הינה $\epsilon - ensemble$ עבור $[n_{in}, k_{in}, d_{in}]$ עבור $\epsilon > 0$ כלשהו, אזי $C^* = [n_{out} \cdot n_{in}, k_{out} \cdot k_{in}, d_{out} \cdot d_{in} \cdot (1-\epsilon)]$.

<u>הגדרה:</u> אוסף קודים $\{C_1, \dots, C_N\}$ נקרא $\epsilon - ensemble$ עבור $[n,k,d]$, אם לפחות $(1-\epsilon)N$ מקודים אלה הינם מהצורה $[n,k,d]$.

<u>משפט:</u> לכל $\epsilon > 0$ קיים k כך שלכל $K \geq k$ ישנם $N = 2^K - 1$ קודים $C_1, \dots, C_N$, כאשר לכל $1 \leq i \leq N$ $C_i: \{0,1\}^K \to \{0,1\}^{2K}$ וגם לפחות $(1-\epsilon)2^K N$ מקודים הנ״ל מקיימים $\mathrm{d(C_i)} \geq h^{-1}\left(\frac{1}{2} - \epsilon\right) 2K = d_{in}$.

<u>הערות:</u>

1. $C_1, \dots, C_N$ הינם קודי Wozencraft Ensemble.
2. אנחנו מקבלים כי $N(1-\epsilon)$ קודים מתוכם ״טובים״ כמו Gillbert-Varshamov.
3. h^{-1} הינה הפונקציה ההופכית של פונקציית האינטרופיה.

<u>הוכחה:</u> נתבונן בשדה F_{2^K} ועל העתקה חחע״ל כלשהי $\sigma: F_{2^K} \to F_2^K$.

לכל $0 \neq \alpha \in F_{2^K}$ נבנה קוד $C_\alpha: F_2^K \to F_2^{2K}$ כך ש $C_\alpha(x) = \left(x, \sigma(\sigma^{-1}(x) \cdot \alpha)\right) =^{(*)} (x, \alpha x)$.

(*) סימון מתמטי לא מדויק.

<u>למה:</u> לכל $\alpha_1 \neq \alpha_2 \in F_{2^K}$ שונים מאפס מתקיים $Im\left(C_{\alpha_1}\right) \cap Im\left(C_{\alpha_2}\right) = \{0\}$.

<u>הוכחה:</u> יהיו $0 \neq x_1, x_2 \in F_2^K$.

אם $x_1 \neq x_2$ אזי $C_{\alpha_1}(x_1) = (x_1, \dots) \neq (x_2, \dots) = C_{\alpha_2}(x_2)$.

אם $0 \neq x_1 = x_2 = x$ אזי $\alpha_1 x \neq \alpha_2 x$ ולכן $C_{\alpha_1}(x_1) = (x, \alpha_1 x) \neq (x, \alpha_2 x) = C_{\alpha_2}(x_2)$.

נבחין כי ניתן לבחור σ כך ש C_α הינו קוד ליניארי.

כעת נראה כי מתוך $C_{\alpha_1}, \dots, C_{\alpha_N}$ ישנם לפחות $(1-\epsilon)2^K$ קודים ״טובים״ המקיימים $\mathrm{d}\left(\mathrm{C}_{\alpha_\mathrm{i}}\right) \geq d_{in}$.

ראשית, נבחין כי מספר הקודים ה״לא טובים״ הינו לכל היותר כמספר המילים ה״לא טובות״.

נסמן $B = \{C_\alpha : d(C_\alpha) < d_{in}\}, S = \{y \in F_2^{2K} : wt(y) < d_{in}\}$.

$$|B| \leq |S| = B(0, d_{in}) = 2^{h\left(\frac{d_{in}}{2K}\right) \cdot 2K} = 2^{h\left(h^{-1}\left(\frac{1}{2} - \epsilon\right)\right)2K} = 2^{\left(\frac{1}{2} - \epsilon\right)2K} = 2^{K - 2\epsilon K} \leq^{(*)} \epsilon \cdot 2^K$$

(*) עבור K מספיק גדול.

מכאן שמתוך $C_{\alpha_1}, \dots, C_{\alpha_N}$ ישנם לפחות $(1-\epsilon)2^K$ קודים ״טובים״ המקיימים $\mathrm{d}\left(\mathrm{C}_{\alpha_\mathrm{i}}\right) \geq d_{in}$.

<u>טענה:</u> האורך של קוד Justesen הינו $2m(2^m-1)$.

<u>הסבר:</u> מכפלה בין האורך של קוד RS לבין האורך של כל אחד מקודי Wozencraft Ensemble.

<u>טענה:</u> המימד של קוד Justesen הינו $m(d+1)$.

<u>הסבר:</u> המימד של קוד Justesen הינו $\log_2|A|$ כאשר $A=\{polynomials\ with\ rank\ d\ over\ F_{2^m}\}$, ומתקיים כי $|A|=(2^m)^{d+1}$. מכאן ש $\log_2|A|=\log_2(2^m)^{d+1}=m(d+1)$.

<u>טענה:</u> המרחק של קוד Justesen הינו $2m\ d\cdot h^{-1}\left(\frac{1}{2}-\epsilon\right)$.

<u>הסבר:</u> המרחק של RS הינו d. המרחק של $(1-\epsilon)2^m$ מתוך קודי Wozencraft Ensemble הינו $2m\ h^{-1}\left(\frac{1}{2}-\epsilon\right)$. מכאן שהמרחק של קוד Justesen הינו $2m\ d\cdot h^{-1}\left(\frac{1}{2}-\epsilon\right)$.

<u>סיכום:</u> עבור $d=\frac{2^m}{2}$ נקבל כי המרחק של קוד Justesen הינו $2^m m\ h^{-1}\left(\frac{1}{2}-\epsilon\right)$ והמימד שלו הינו $m2^{m-1}$.
מכיוון שאורך הקוד הינו $2m(2^m-1)$, נקבל כי $\delta=\frac{h^{-1}\left(\frac{1}{2}-\epsilon\right)}{2}$, $R=\frac{1}{4}$. כלומר, קודי Justesen הינם קבוצה של קודים "טובים" בעלי $R,\delta>0$.