# קוונטיזציה סקלרית ווקטורית תחת שגיאות-ערוץ

מאת

גל בן-דוד

# קוונטיזציה סקלרית ווקטורית תחת שגיאות-ערוץ

חיבור על מחקר
לשם מילוי חלקי של הדרישות לקבלת תואר
דוקטור למדעים

מאת

גל בן-דוד



הוגש לסנט הטכניון - מכון טכנולוגי לישראל

סיון תשנ״ה חיפה יוני 1995

המחקר נערך בהנחייתו של פרופסור דוד מלאך
בפקולטה להנדסת חשמל.

תודתי העמוקה נתונה לפרופי דוד מלאך על הצעת הנושא, הכוונתו והנחייתו המסורה.

חיבור זה מהווה סיום של תקופה נהדרת, שהחלה עוד בביה״ס התיכון, בה רכשתי את השכלתי המקצועית במעבדה לעיבוד אותות.
תודתי העמוקה נתונה ליורם אור-חן על הפרויקט הראשון במעבדה ועל התמיכה לאורך השנים.
תודה כמובן גם לזיוה, נמרוד, אורי, אבי ותמרה על התקופה הנהדרת.

ברצוני להודות לקרן וולף על התמיכה הכספית הנדיבה במסגרת המלגה לזכרו של ד״ר משה גלבוע ז״ל.

לרנטה ודן-אילן היקרים

# תוכן העניינים

עמוד

## תוכן העניינים (המשך)

# תוכן העניינים (המשך)

# תקציר


למערכות קוונטיזציה וקטורית (וקוונטיזציה סקלרית כמקרה פרטי) קיים שימוש נרחב ביישומים של קידוד מקור. ניתן למצוא מערכות קוונטיזציה וקטורית רבות ביישומים של קידוד דיבור ותמונה. קוונטייזר וקטורי מחלק את קבוצת כל האותות (הוקטורים) האפשריים למספר סופי של תאי-חלוקה, כאשר כל תא חלוקה מיוצג ע"י וקטור ייצוג. קוונטיזציה וקטורית מהווה לכן שיטה לדחיסה עם עיוות. כלומר, בתהליך אובד חלק מן המידע הקיים באות והשחזור שלו איננו זהה למקור. ההבדל בין האות המשוחזר לאות המקורי נקרא עיוות הקוונטיזציה.

מערכת תקשורת המבוססת על קוונטייזר וקטורי בנויה משני חלקים: מקודד ומפענח. המקודד (לפני הכניסה לערוץ) בודק לאיזה תא חלוקה שייך הוקטור אותו עליו לקודד. גם במקודד וגם במפענח קיימת טבלה של וקטורי הייצוג, כאשר לכל וקטור ייצוג משובץ אינדקס ספרתי. האינדקס הספרתי משודר בערוץ והמפענח משחזר את וקטור הייצוג מן הטבלה.

תכן קוונטייזר וקטורי מבוסס על ידיעת הסטטיסטיקה של מקור האות, כאשר מעשית נעשה התכן על סמך סדרת אימון שנדגמה במוצא מקור האות. שיטות הניתוח והתכן "הקלסיות" של מערכות קוונטיזציה מניחות כי אף על פי שהמידע הספרתי עובר בערוץ תקשורת, לא נגרמים שיבושים בתהליך העברתו. בערוצי תקשורת מעשיים נעשה שימוש בקידוד-ערוץ אשר במקרים רבים מביא את השפעת שגיאות הערוץ להיות זניחה.

במקרים רבים אחרים השפעת שגיאות הערוץ איננה זניחה, ושימוש בקוונטייזרים אשר לוקחים בחשבון את השגיאות בערוץ יביאו לביצועים טובים יותר.

אחת המטרות של עבודה זו היא לבחון את השפעת שגיאות הערוץ על ביצועי מערכות קוונטיזציה וקטורית נתונות. השגיאות בערוץ התקשורת גורמות לאינדקס שגוי להופיע במפענח, אשר יפוענח לוקטור שחזור שגוי. המרחק בין הוקטור שנתקבל לוקטור שאמור היה להתקבל במוצא המפענח נקרא עיוות הערוץ. המרחק בין הוקטור (השגוי) שנתקבל לאות המקורי נקרא העיוות הכולל. בעבודה זו אנו מניחים כי ערוץ התקשורת הינו <u>חסר-זיכרון</u>, כאשר חלק מן התוצאות נתקבלו, בפרט, עבור הערוץ הבינארי הסימטרי.

בנוסף למציאת חסמים על ביצועי מערכות קוונטיזציה וקטורית נתונות, העבודה מתרכזת בשני גורמים המשפיעים על עיוות הערוץ ונתונים בידי המתכנן. הגורם הראשון הוא מבנה הקוונטייזר הוקטורי (והסקלרי) מבחינת תאי חלוקה ווקטורי ייצוג. הגורם השני הוא שיבוץ האינדקסים המשודרים בערוץ לוקטורי הייצוג.

בחלקה הראשון של העבודה מפותחים חסמים על עיוות הערוץ <u>מעל כל מקורות האות האפשריים</u>. חסמים אלו מספקים למתכנן מידע כמותי על עיוות הערוץ. מוצגות מספר דוגמאות מספריות להדגמת החסמים. עבור המקרה המיוחד של קוונטייזר סקלרי אחיד, קוד בינארי טבעי וערוץ בינארי סימטרי, החסם העליון מתלכד מעשית עם החסם התחתון. בנוסף, עיוות הערוץ עבור המקרה הזה איננו תלוי מעשית במספר הסיביות של הקוונטייזר. בדוגמאות אחרות, המרווח בין החסם העליון לחסם התחתון הוא גדול יחסית וידיעתם איננה תורמת מספיק להערכת ביצועי המערכת. עבור מקרים אלו מוצגים חסמים הדוקים יותר עבור משפחה מצומצמת יותר של מקורות, המאופיינת ע"י מומנטים של הסתברות המקור ובפרט אילוץ על

הספק המקור. כמו-כן מוצגים חסמים אסימפטוטיים עבור ערוץ בינארי סימטרי (בלי או עם קוד לתיקון שגיאות) כאשר הסתברות השגיאה בערוץ שואפת לאפס.

עבור קוונטייזר סקלרי אחיד בן 4 סיביות ויותר, עיוות הערוץ עבור השיבוץ הבינארי הטבעי קטן יותר מאשר ממוצע העיוות מעל כל השיבוצים האפשריים, עבור כל מקור שהוא.

בעבודות קודמות שנעשו בנושא הוצגו תנאים הכרחיים לאופטימליות של קוונטייזר וקטורי בתנאי שגיאות ערוץ. תנאים אלו מביאים לתהליך תכן מסובך, אשר מניב ספרייה של קוונטיזרים שונים, שכל אחד מהם מתאים לתנאי ערוץ שונים. שמירת ספרייה כזו במפענח ובמקודד עולה בדרישות זיכרון יקרות. בעבודה זו אנו מציעים שיטה תת-אופטימלית הדורשת רק כיוון ליניארי של תאי-החלוקה ווקטורי הייצוג בתלות בשגיאות הערוץ. תהליך התכן פשוט הרבה יותר, ומזדהה ברובו עם התכן ה-"קלסי" שלא מתחשב בשגיאות ערוץ. גם הדרישות על הזיכרון במערכת קטנות בהרבה. בעבודה מוצגות דוגמאות מספריות המראות שהתשלום בביצועים, יחסית לקוונטייזר המתוכנן לפי התנאים לאופטימליות, הוא נמוך.

גורם חשוב המשפיע על עיוות הערוץ הוא שיבוץ האינדקסים לוקטורי הייצוג. כללית, המטרה היא לגרום לשגיאות ערוץ לגרום לנזק הקטן ביותר האפשרי. ישנו מספר עצום של שיבוצים אפשריים (לקוונטייזר בעל $N$ וקטורי ייצוג יש $N!$ שיבוצים אפשריים) שבד"כ אינו מאפשר מעשית את הבדיקה של כולם.

עבור המקרה המיוחד של קוונטייזר סקלרי אחיד המוזן ממקור אחיד, ניתנת הוכחה כי השיבוץ הבינארי הטבעי הוא שיבוץ אופטימלי, המביא לעיוות הערוץ הקטן ביותר.

בהמשך מפותחים חסמים על עיוות הערוץ, מעל כל השיבוצים האפשריים, עבור ערוצים חסרי-זיכרון כלשהם. מוצגות דוגמאות מספריות המשוות את החסמים עם כעשרת אלפים שיבוצים שנבחרו באקראי.

בחלקה האחרון של העבודה מפותח אלגוריתם לשיבוץ תת-אופטימלי של אינדקסים לוקטורי הייצוג בקוונטייזר וקטורי, הפועל תחת ערוץ סימטרי חסר זיכרון. אנו מראים כי מימוש האלגוריתם הוא בסיבוכיות של $O\left(N^3\right)$.

בדוגמאות מספריות, הביצועים המתקבלים מן האלגוריתם טובים מן הביצועים המתקבלים מאלגוריתם תת-אופטימלי נפוץ בספרות של חילופי אינדקסים, שהוא בסיבוכיות $O\left(N^4\right)$.

בנוסף, מוצגת גם השוואה בין החסמים על עיוות הערוץ מעל כל השיבוצים האפשריים לבין חיפוש בעשרת אלפים שיבוצים אקראיים ועבור שני מקרים נערכת השוואה גם עם השיבוץ התת-אופטימלי המוצע. מן ההשוואות עולה כי החיפוש האקראי איננו מביא לתוצאות מספקות בגלל העושר הרב של שיבוצים אפשריים (עבור קוונטייזר של 4 סיביות בלבד קיימים $2\cdot10^{13}$ שיבוצים אפשריים). עבור המקרה המיוחד של קוונטייזר סקלרי אחיד ומקור אחיד, מתלכד השיבוץ המוצע עם החסם התחתון. במקרה טיפוסי שנבדק בסימולציה, נמוך עיוות הערוץ עבור השיבוץ המוצע ב-8dB בהשוואה לממוצע של עיוות הערוץ מעל כל השיבוצים האפשריים. עבור אותו מקרה, העיוות כתוצאה מאלגוריתם חילופי האינדקסים גבוה ב-1dB מן העיוות עבור השיבוץ המתקבל מן האלגוריתם המוצע.

# רשימת סמלים וקיצורים

$\lfloor x \rfloor$ - הערך השלם של $x$

$\delta_{ij}$ - דלתה של Kronecker

$K$ - מימד הוקטורים אותם מקודד הקוונטייזר הוקטורי

$N$ - מספר וקטורי/רמות הייצוג של קוונטייזר וקטורי/סקלרי

$L$ - מספר הסיביות של קוונטייזר וקטורי/סקלרי

$L^0$ - החזקה הדומיננטית של שגיאות הערוץ בערוץ בינארי סימטרי

$R^Q$ - קצב פליטת המידע במוצא המקודד (בסיביות לדגימה)

$R_i$ - תא/אינטרוול החלוקה ה- $i$ בקוונטייזר וקטורי/סקלרי

$f_{\underline{x}}$ - צפיפות ההסתברות של המשתנה האקראי הוקטורי $\underline{x}$

$\mathcal{Q}$ - אופרטור הקוונטייזר הוקטורי (כולל מקודד ומפענח)

$\mathcal{E}$ - אופרטור הקידוד

$\mathcal{D}$ - אופרטור הפיענוח

$D_{\text{Total}}$ - עיוות כולל בקוונטייזר וקטורי/סקלרי

$D_Q$ - עיוות הקוונטיזציה בקוונטייזר וקטורי/סקלרי

$D_c$ - עיוות הערוץ בקוונטייזר וקטורי/סקלרי

$\overline{D}_c$ - ממוצע עיוות הערוץ בקוונטייזר וקטורי/סקלרי על פני כל השיבוצים האפשריים

$\underline{0}$ - וקטור $N$-מימדי המורכב כולו מאפסים

$\underline{1}$ - וקטור $N$-מימדי המורכב כולו מאחדים

$\underline{\rho}$ - וקטור רמות הייצוג בקוונטייזר סקלרי אחיד

$\rho_i$ - רמת הייצוג ה- $i$ בקוונטייזר סקלרי אחיד

$\underline{s}_i$ - אות בערוץ תקשורת

$\underline{z}$ - וקטור כללי $N$-מימדי

$C_i$ - מטריצה המכילה אחדים בעמודה ה- $i$ והיתר אפסים

$\underline{\phi}_i$ - וקטור הייצוג ה- $i$ של קוונטייזר וקטורי

$d(\cdot,\cdot)$ - מידת מרחק בין וקטורים

$D$ - מטריצת המרחקים בין וקטורי הייצוג של קוונטייזר וקטורי/סקלרי

$\hat{D}$ - מטריצת מרחקים משוקללת בין וקטורי הייצוג של קוונטייזר וקטורי/סקלרי

$\tilde{D}$ - מטריצת מרחקים משוקללת מנורמלת

$Z$ - מטריצה יוניטרית המכילה את הוקטורים העצמיים של המטריצה $\tilde{D}$

$\underline{z}_i$ - הוקטור העצמי ה- $i$ של המטריצה $\tilde{D}$

$\Omega$ - מטריצה אלכסונית המכילה את הערכים העצמיים של המטריצה $\tilde{D}$

$\omega_i$ - הערך העצמי ה- $i$ של המטריצה $\tilde{D}$

$S_i$ - סכום האיברים בשורה ה- $i$ של מטריצת מרחקים משוקללת

$S$ - סכום משתני עזר הנוספים למטריצת מרחקים משוקללת

$\Delta$ - מרחק בין רמות ייצוג שכנות בקוונטייזר סקלרי אחיד

$P$ - מטריצה אלכסונית המכילה את הסתברויות המקור

$Q$ - מטריצת הסתברויות הערוץ

$Q_N$ - מטריצת הסתברויות הערוץ במקרה של ערוץ בינארי סימטרי ו- $N$ סיביות

$Q_H$ - מטריצת הסתברויות הערוץ עבור ערוץ בינארי סימטרי עם קוד Hamming לתיקון שגיאות

$c(i)$ - מלת הקוד ה- $i$ עבור קוד לתיקון שגיאות

$H(i,j)$ - מרחק Hamming בין הייצוגים הבינאריים של המספרים $i$ ו- $j$

$q$ - הסתברות שגיאה בערוץ בינארי סימטרי

$V$ - מטריצה יוניטרית המכילה את הוקטורים העצמיים של מטריצת הערוץ $Q$

$\underline{v}_i$ - הוקטור העצמי ה- $i$ של מטריצת הערוץ $Q$

$U$ - מטריצה יוניטרית שעמודתה הראשונה היא $\underline{1}$

$\Lambda$ - מטריצה אלכסונית המכילה את הערכים העצמיים של מטריצת הערוץ $Q$

$\lambda_i$ - הערך העצמי ה- $i$ של מטריצת הערוץ $Q$

$\Pi$ - אופרטור פרמוטציה

$\pi$ - מטריצת פרמוטציה

$L(\cdot)$ - אופרטור לגרנג'

$\mu_i$ - כופל לגרנג' ה- $i$

BER - Bit Error Rate

BSC - Binary Symmetric Channel

COVQ - Channel Optimized Vector Quantizer

LP - Linear Programming

LPC - Linear Predictive Coding

NBC - Natural Binary Code

NVQ - Noiseless Vector Quantizer

PDF - Probability Density Function

QA - Quadratic Assignment

RVQ - Robust Vector Quantizer

SA - Simulated Annealing

SQ - Scalar Quantizer

SVQ - Scaled Vector Quantizer

VAPC - Vector Adaptive Predictive Coding

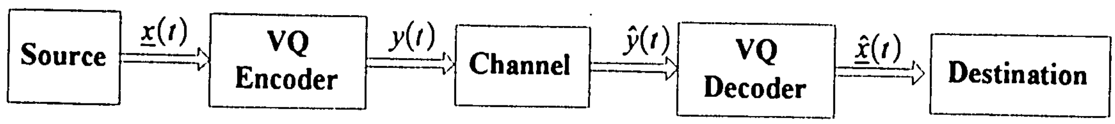
VQ - Vector Quantizer

# פרק 1 - מבוא

## 1.1 תיאור הנושא

עבודת מחקר זו עוסקת בחקר הביצועים של מערכות קידוד מסוג קוונטיזציה וקטורית כאשר המידע הספרתי מועבר דרך ערוצי תקשורת רועשים. בנוסף מוצעות במחקר דרכים לשיפור הביצועים של מערכות אלו.

מטרתן של מערכות **קוונטיזציה וקטורית** (VQ - Vector Quantization) **וקוונטיזציה סקלרית** (SQ - Scalar Quantization) היא ייצוג אותות אנלוגיים בצורה ספרתית בנאמנות הגבוהה ביותר שניתן להשיג במגבלות קצב (בסיביות לדגימה) וסיבוכיות (בפעולות חישוביות לדגימה). את המידע הספרתי ניתן לשדר דרך ערוצים ספרתיים. דוגמאות רבות לשימוש בקוונטיזציה וקטורית וסקלרית לקידוד אותות דיבור ותמונות ניתן למצוא ב-[Ge92] וב-[Ja84]. בציור 1.1 מתוארת באופן סכימטי מערכת תקשורת מבוססת VQ.




ציור 1.1 - מערכת קוונטיזציה וקטורית

Fig. 1.1 - Vector Quantization System

מערכת התקשורת מורכבת **ממקודד** (Encoder) **ומפענח** (Decoder). המקודד מקבל בזמן $t$ **ממקור האות** (Source) וקטור $\underline{x}(t)$ של $K$ דגימות אותו על המערכת לקודד, ולהעביר ליעד (Destination) דרך ערוץ ספרתי (Channel). עבור המקרה הפרטי $K=1$, מערכת הקידוד נקראת קוונטיזציה סקלרית - SQ. מערכת ה-VQ מקודדת את וקטור הכניסה **למלת קוד** (Codeword) ספרתית $y(t)$ המשודרת בערוץ.

מלת הקוד הספרתית $y(t)$ נבחרת מתוך **אלפבית הערוץ** (Channel Alphabet) הכולל מספר סופי של ערכים אפשריים. הערוץ איננו מושלם ולעתים קורות בו שגיאות. מסיבה זו המידע הספרתי במוצא הערוץ $\hat{y}(t)$ איננו זהה לחלוטין למידע הנכנס אליו. המפענח מקבל את המידע במוצא הערוץ ומפיק וקטור ייצוג/שחזור $\hat{\underline{x}}(t)$ (Representation/Reconstruction Vector) בן $K$ דגימות המתאים ל-$\hat{y}(t)$ מתוך טבלת וקטורי הייצוג (Codebook). וקטור הייצוג מסופק ליעד (Destination).

גם מימוש **המקודד** נעשה לרוב בעזרת אותה טבלת וקטורי ייצוג. עם קבלת וקטור $\underline{x}(t)$ לקידוד, המקודד עורך חיפוש בין וקטורי הייצוג למציאת המועמד הקרוב ביותר אליו (לפי תנאי השכן הקרוב שיתואר בפרק 2).

כאמור, בערוץ משודר אינדקס ספרתי המצביע על וקטור ייצוג. בנוכחות שגיאות ערוץ ישנה חשיבות באיזה שיבוץ אינדקסים ספציפי נעשה שימוש.

אין אפשרות להעביר את הוקטור $\underline{x}(t)$ במדויק ליעד משתי סיבות עיקריות:

א. האות הפיזיקלי המקורי הוא לרוב אות אנלוגי בעל רצף של ערכים אפשריים. הערוץ הספרתי מסוגל להעביר ערכים מתוך אלפבית הערוץ שמספרם סופי. המפענח יפיק לכן מספר סופי של וקטורי שחזור אפשריים. ולכן, גם עבור ערוץ מושלם (שעבורו המוצא הספרתי זהה לכניסה $\hat{y}(t) = y(t)$ בכל עת) קיים עיוות בין וקטור הכניסה לוקטור היציאה. עיוות זה נקרא **עיוות הקוונטיזציה** (Quantization Distortion).

במציאות נדגם האות ע"י ממיר אנלוגי לספרתי סקלרי כשלב ראשון (Analog to Digital Conversion) והמידע הנכנס למקודד הוא ספרתי. במקרה זה עיוות הקוונטיזציה נובע מן השימוש ב-VQ (או SQ) לדחיסת המידע. מספר הערכים האפשריים בוקטור הכניסה למקודד גדול ממספר הערכים האפשריים באלפבית הערוץ.

ב. שגיאה בערוץ תגרום לכך שהמפענח יספק ליעד וקטור שחזור מוטעה. העיוות הנגרם כתוצאה משגיאה זו נקרא **עיוות הערוץ** (Channel Distortion).

עבודה רבה מדווחת בספרות בנושא של הקטנת עיוות הקוונטיזציה עבור ערוצים מושלמים (ללא עיוות ערוץ). דוגמאות והפניות למאמרים נתן למצוא בספרים [Ge92] ו-[Ja84]. עבודות רבות נוספות עוסקות במימוש יעיל של מערכות VQ, ללא עיוות ערוץ, הן בשלבי התכן והן במימוש. במקרים מעשיים כאשר קיים ערוץ לא-מושלם ניתן להוסיף קוד לתיקון שגיאות (Error Correcting Code). במקרים רבים, שימוש בקוד לתיקון שגיאות מספק למערכת הקוונטיזציה ערוץ שקול (הכולל מקודד ערוץ, ערוץ ומפענח ערוץ) עם הסתברויות שגיאה נמוכות מאד, כך שמעשית ניתן להזניח את עיוות הערוץ.

עבודה זו עוסקת באותם מקרים, שאף הם רבים, בהם, משיקולי מימוש, לא משתמשים בקוד לתיקון שגיאות או שהקוד איננו מספק ערוץ שקול עם הסתברויות שגיאה נמוכות מספיק. במקרים אלו, נוצרות שגיאות במידע הספרתי ועיוות הערוץ הוא גורם שיש להתחשב בו. הסעיף הבא מתאר את מטרות עבודת המחקר ואת תוצאותיה העיקריות ביתר פירוט.

## 1.2 מטרות העבודה ותוצאות עיקריות

מטרת עבודת מחקר זו היא חקר עיוות הערוץ ומציאת חסמים על גודלו במערכות קוונטיזציה סקלרית ווקטורית. מוצעות בעבודה גם מספר דרכים לשיפור הביצועים של קוונטייזרים סקלריים ווקטוריים הפועלים תחת שגיאות ערוץ.

התוצאות העיקריות שהושגו בעבודה זו הן:

א. על מנת לקבל מידע כמותי על עיוות הערוץ במערכות קיימות פותחו חסמים על עיוות הערוץ מעל כל מקורות האות האפשריים. העבודה מציגה דרך למציאת חסמים, עליון ותחתון, עבור ערוצים חסרי-זיכרון כלשהם. כמו-כן, מוצגים חסמים אסימפטוטיים עבור ערוץ בינארי סימטרי (בלי או עם קוד לתיקון שגיאות) כאשר הסתברות השגיאה בערוץ שואפת לאפס.

ב. עבודות רבות בספרות דנות בשינוי מבנה הקוונטייזר הוקטורי כך שביצועיו ישתפרו תחת שגיאות ערוץ. הפתרונות האופטימליים המוצעים הנם יקרים ומורכבים מבחינת התכן והמימוש. עבור הערוץ הבינארי הסימטרי, אנו מציעים שינוי פשוט במבנה הקוונטייזר המביא לשיפור (תת-אופטימלי) של הביצועים. שיטה זו הנה פשוטה לתכן וזולה למימוש, עם ביצועים הפחותיים אך במעט בהשוואה לאלגוריתמים האופטימליים.

ג. עיוות הערוץ תלוי במידה רבה בשיבוץ מילות הקוד לוקטורי הייצוג. בעבודה מוצגים חסמים על עיוות הערוץ מעל כל השיבוצים האפשריים של מלות קוד לערוצים חסרי זיכרון. חסמים אלו מאפשרים למתכנן להעריך את השפעת השיבוץ על המערכת שבתכנון.

ד. עבור ערוץ סימטרי חסר זיכרון מוצע אלגוריתם תת-אופטימלי לשיבוץ של מילות קוד לוקטורי הייצוג בקוונטייזר. האלגוריתם דורש פחות חישובים מאלגוריתם תת-אופטימלי אחר המקובל בספרות. בדוגמאות מספריות מניב האלגוריתם המוצע שיבוץ עדיף על האלגוריתם המקובל.

בנוסף לכך, עבור המקרה המיוחד של קוונטייזר סקלרי אחיד ומקור אחיד ניתנת הוכחה כי הקוד הבינארי הטבעי הוא השיבוץ האופטימלי. בעבודה מוצגות תכונות נוספות של מקרה זה (גם בהקשר של סעיף א' לעיל).

## 1.3 מבנה העבודה

בפרק 2 ניתן רקע כללי ותיאור מערכות קוונטיזציה הפועלות תחת שגיאות ערוץ. כמו-כן נסקרים מקורות הספרות הדנים במערכות אלו ודרכים שהוצעו לשיפור ביצועי הקוונטייזרים תחת שגיאות ערוץ. פרק 3 עוסק בחסמים על עיוות הערוץ מעל כל מקורות האות האפשריים (סעיף א' לעיל). פרק 4 דן בכיוונון ליניארי של קוונטייזרים וקטורים לשיפור הביצועים תחת שגיאות ערוץ (סעיף ב'). פרק 5 דן בהוכחת האופטימליות של השיבוץ הבינארי הטבעי עבור קוונטייזר סקלרי אחיד ומקור אחיד (סעיף ד'). פרק 6 עוסק במציאת חסמים על עיוות הערוץ מעל כל השיבוצים האפשריים עבור ערוצים חסרי זיכרון (סעיף ג'). בפרק 7 מתואר האלגוריתם המוצע לשיבוץ של מלות הקוד (סעיף ד'). סיכום והצעות להמשך מחקר מופיעים בפרק 8.

# פרק 2 - רקע וסקר ספרות

## 2.1 קוונטיזציה וקטורית - מבוא

קוונטיזציה וקטורית תופסת מקום נכבד במערכות **קידוד מקור** (Source Encoding) מודרניות. הקוונטיזציה הסקלרית, מקרה פרטי של הקוונטיזציה הוקטורית, שימושית עדיין במערכות רבות בזכות פשטות המימוש והתכן. בנוסף לכך, הקוונטיזציה הסקלרית רגישה פחות לחוסר התאמה בין המקור עבורו היא תוכננה למקור במציאות. דוגמאות רבות והפניות למאמרים העוסקים בקוונטיזציה וקטורית וסקלרית לקידוד אותות דיבור ותמונה נתן למצוא ב-[Ge92] וב-[Ja84]. המוטיבציה לשימוש בקוונטייזרים וקטורים נובעת מתורת האינפורמציה [Vi79]. עבור קצב שידור (Transmission Rate) נתון בערוץ - $R$, העיוות (Distortion) המתקבל מקוונטייזר וקטורי יהיה קטן יותר מהעיוות הנגרם ע"י קוונטייזר סקלרי.

קוונטייזר וקטורי $Q$ במימד (Dimension) $K$ ובגודל (Size) $N$ הוא מיפוי מהמרחב האוקלידי ה- $K$ מימדי, $\Re^K$, לקבוצה סופית $C$ המכילה $N$ איברים $i=0,1,\ldots,N-1$ , $\underline{\phi}_i$, המכונים וקטורי ייצוג או וקטורי שחזור (Representation/Reconstruction Vectors). וקטורי הייצוג הנם גם כן ב-$\Re^K$. הקבוצה $C$ מכונה ספר הקוד (Codebook).

$$Q:\Re^K \rightarrow C=\left(\underline{\phi}_0,\underline{\phi}_1,\ldots,\underline{\phi}_{N-1}\right) \tag{2.1}$$

השגיאה או העיוות הנוצר מפעולת הקוונטיזציה הינו המרחק בין וקטור הכניסה $\underline{x}$ לוקטור הייצוג המתקבל:

$$d\left(\underline{x},Q(\underline{x})\right) \tag{2.2}$$

קצב הקוונטייזר (Quantizer Rate) מוגדר ע"י: $\dfrac{\text{Bits}}{\text{Sample}}$ $R^Q=\dfrac{\log_2 N}{K}$.

הקוונטייזר מבצע חלוקה (Partition) של מרחב הכניסה $\Re^K$ ל- $N$ תאי חלוקה (Cells) זרים. לכל וקטור ייצוג קיים תא חלוקה כך שכל הוקטורים בתא זה מיוצגים ע"י וקטור הייצוג. תא החלוקה ה- $i$:

$$R_i=\left\{\underline{x}\in\Re^K : Q(\underline{x})=\underline{\phi}_i\right\} \tag{2.3}$$

תאי החלוקה הנם זרים זה לזה ומכסים יחדיו את כל המרחב:

$$\bigcup_i R_i=\Re^K\,,\quad R_i\cap R_j=\varnothing \text{ for } i\neq j \tag{2.4}$$

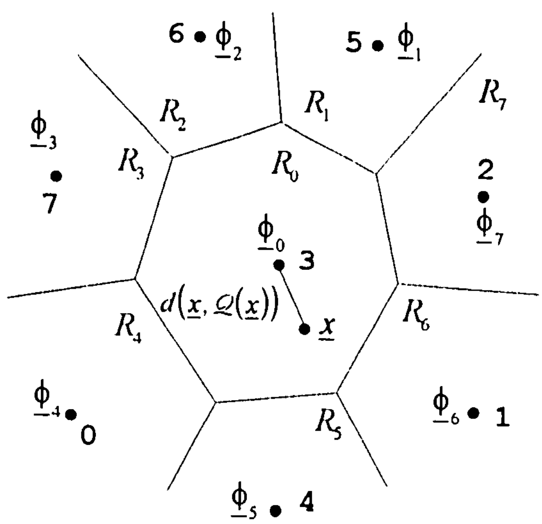
תא-חלוקה יכול להיות חסום (Bounded) ואז הוא נקרא תא "גרגירי" (Granular Cell) או בלתי חסום ואז הוא נקרא תא "העמסה" (Overload Cell).

**<u>הגדרה</u>**: קוונטייזר וקטורי נקרא <u>רגולרי</u> אם תאי החלוקה הנם קמורים (Convex) וכל וקטור ייצוג נמצא בתא אותו הוא מייצג - $\forall i \quad \underline{\phi}_i \in R_i$. בהמשך נראה כי עבור מידת מרחק ריבועית, קוונטייזר המקיים את תנאי האופטימליות לתכן קוונטייזר ללא שגיאות ערוץ הוא רגולרי. תחת שגיאות ערוץ תכונה זו איננה מובטחת.

ניתן לפצל את פעולת הקוונטייזר הוקטורי לשני רכיבים: המקודד (Encoder) והמפענח (Decoder). המקודד $\mathcal{E}$ הוא מיפוי מוקטורי הכניסה לקבוצת אינדקסים $I = (0,1,\ldots,N-1)$ והמפענח $\mathcal{D}$ הוא מיפוי מקבוצת האינדקסים לספר הקוד $C$.

$$\begin{aligned} &\mathcal{E} : \Re^K \to I \quad \text{and} \quad \mathcal{D} : I \to C \\ &\mathcal{Q}(\underline{x}) = \mathcal{D}(\mathcal{E}(\underline{x})) \end{aligned} \tag{2.5}$$

האינדקסים מן הקבוצה $I$ הם המשודרים בערוץ.
אין קביעה חד-משמעית של השיבוץ (Index Assignment) בין הוקטורים המייצגים לאינדקסים המתאימים להם וניתן לבצע $N!$ שיבוצים אפשריים. השיבוץ מתואר ע"י מטריצת פרמוטציה $\pi$, המכילה רק אפסים ואחדים וסכום כל שורה ועמודה שלה שווה לאחד.
בציור 2.1 מתואר סכמטית קוונטייזר וקטורי רגולרי:




ציור 2.1 - תאור סכימטי של קוונטייזר וקטורי רגולרי

Fig. 2.1 - Schematic drawing of a Regular Vector Quantizer

ניתן לראות בסכמה את אזורי החלוקה $R_i$, את וקטורי הייצוג $\underline{\phi}_i$ ואת האינדקס המותאם לכל וקטור ייצוג.

השיבוץ שנבחר בדוגמה הוא:

$$\begin{matrix} \{\underline{\phi}_0, \underline{\phi}_1, \underline{\phi}_2, \underline{\phi}_3, \underline{\phi}_4, \underline{\phi}_5, \underline{\phi}_6, \underline{\phi}_7\} \\ \{3\ ,\ 5,\ 6,\ 7,\ 0, 4,\ 1,\ 2\} \end{matrix} \tag{2.6}$$

לדוגמה, וקטור הכניסה המסומן ב-$\underline{x}$ נכלל בתוך $R_0$ ולכן מיוצג ע"י $\underline{\phi}_0$ והאינדקס 3 ישודר בערוץ.

המרחק הממוצע בין וקטור הכניסה לקוונטייזר לוקטור היציאה (ללא שגיאות ערוץ), נקרא עיוות הקוונטיזציה (Quantization Error).

$$\begin{aligned} D_Q &= E\left[d\left(\underline{x}, Q(\underline{x})\right)\right] = \\ &= \int_{\Re^K} d\left(\underline{x}, Q(\underline{x})\right) \cdot f_{\underline{x}}(\underline{x}) \cdot d\underline{x} = \\ &= \sum_{i=0}^{N-1} \int_{R_i} d\left(\underline{x}, \underline{\phi}_i\right) \cdot f_{\underline{x}}(\underline{x}) \cdot d\underline{x} \end{aligned} \tag{2.7}$$

### 2.1.1 התנאים לאופטימליות של קוונטייזר וקטורי - תכן הקוונטייזר

מטרת התכן של קוונטייזר וקטורי (ללא התחשבות בשגיאות ערוץ) היא מינימום של עיוות הקוונטיזציה. קיימים שני תנאים הכרחיים למינימום של העיוות [Ge92, סעיף 10.1]:

**<u>תנאי השכן הקרוב (Nearest Neighbor Condition)</u>**

בהינתן ספר קוד $C$, וקטור כניסה $\underline{x}$ ישוחזר ע"י וקטור הייצוג הקרוב אליו ביותר:

$$Q(\underline{x}) = \underline{y}_i \quad \text{if } d\left(\underline{x}, \underline{\phi}_i\right) \le d\left(\underline{x}, \underline{\phi}_j\right) \quad \text{for } j = 0, 1, \ldots, N-1 \tag{2.8}$$

תנאי זה קובע את תאי החלוקה $R_i$ , $i = 0, 1, \ldots, N-1$. הבדיקה לאיזה תא שייך וקטור כניסה $\underline{x}$ נעשית ע"י מדידת המרחק ממנו לכל וקטורי הייצוג ומציאת הקרוב אליו ביותר. בנוסף, עבור מידת מרחק אוקלידית, תנאי השכן הקרוב מביא לכך שתא חלוקה הוא חיתוך של חצאי-מרחב (Half Space) המתוארים ע"י על-מישורים (Hyperplane). תא חלוקה כזה הוא קמור (Convex).

**תנאי הצנטרואיד (Centroid Condition)**

בהינתן חלוקה המרחב לתאי חלוקה, וקטור ייצוג $\underline{\phi}_i$ המייצג תא חלוקה $R_i$ הוא הוקטור המביא למינימום העיוות בתא זה:

$$
\begin{aligned}
&Q(\underline{x}) = \underline{\phi}_i \quad \text{for all } \underline{x} \in R_i \\
&\text{if } \mathrm{E}\left\{d(\underline{x},\underline{\phi}_i)\,|\,\underline{x} \in R_i\right\} \le \mathrm{E}\left\{d(\underline{x},\underline{z})\,|\,\underline{x} \in R_i\right\} \quad \text{for all } \underline{z} \in R_i
\end{aligned}
\tag{2.9}
$$

עבור מידת מרחק ריבועית $d(\underline{x},\underline{y}) = \|\underline{x}-\underline{y}\|^2$ תנאי הצנטרואיד מצטמצם לכדי:

$$
\begin{aligned}
&Q(\underline{x}) = \underline{\phi}_i \quad \text{for all } \underline{x} \in R_i \\
&\text{if } \underline{\phi}_i = \mathrm{E}\left\{\underline{x}\,|\,\underline{x} \in R_i\right\} = \frac{\int_{R_i} \underline{x}\cdot f_{\underline{x}}(\underline{x})\cdot d\underline{x}}{\int_{R_i} f_{\underline{x}}(\underline{x})\cdot d\underline{x}}
\end{aligned}
\tag{2.10}
$$

תנאי זה מבטיח כי $\forall i \quad \underline{\phi}_i \in R_i$ וביחד עם תכונת הקמירות, מתאר קוונטייזר רגולרי.

כאמור, תנאים אלו הם תנאים הכרחיים בלבד. תכן קוונטייזר וקטורי נעשה לרב בחישוב איטרטיבי המיישם את שני התנאים לסירוגין. אלגוריתם איטרטיבי כזה המבוסס על ידיעת צפיפות הסתברות של מקור האות $f_{\underline{x}}(\underline{x})$ נקרא אלגוריתם Lloyd והוא מתואר ב-[Ll82].

במקרים מעשיים לא ידועה פונקצית ההסתברות, ואלגוריתם התכן מבוסס על סדרת אימון של וקטורים. אלגוריתם התכן במקרה זה נקרא LBG והוא מתואר ב-[Li80].

קיימות גרסות רבות לאלגוריתם ה-LBG שנועדו בעיקר להקטין את סיבוכיות החישוב בפעולת הקידוד. גרסה אחת כזו נקראת LBG <u>במבנה עץ</u> [Ge92 סעיף 12.4]. גישה זו נותנת ביצועים פחותים מבחינת עיוות קוונטיזציה בהשוואה ל-LBG רגיל, אך היא יעילה במספר חישובים במימוש. תכן ה-VQ במבנה עץ מבוסס על פיצול תאי החלוקה הקיימים בכל איטרציה כאשר הפיצול נעשה רק לפי וקטורי האימון השייכים לתא החלוקה.

**אלגוריתם ה-LBG במבנה עץ[Ge92 סעיף 12.4]**

1. בנה קוונטייזר בן סיבית אחת עם שני תאי חלוקה $R_0^{(1)}, R_1^{(1)}$. הסימון העילי מסמן את מספר הסיביות בקוונטייזר $l \leftarrow 1$. קוונטייזר זה חילק את סדרת האימון לשתי תת-סדרות: וקטורי-אימון השייכים ל-$R_0^{(1)}$ ווקטורי-אימון השייכים ל-$R_1^{(1)}$.

2. פצל כל תא חלוקה לשני תתי-תאים: $R_i^{(l)}$ מפוצל ל-$R_{2i+1}^{(l+1)}, R_{2i}^{(l+1)}$, עבור $i = 0,1,\ldots,2^l - 1$.
**בתהליך פיצול תא חלוקה נעשה שימוש רק בוקטורי-האימון השייכים לאותו תא חלוקה.**

3. כעת הקוונטייזר הוא בעל סיבית נוספת $l \leftarrow l+1$.
אם הגעת למספר הסיביות הדרוש $l = L$ סיים, אחרת חזור לצעד 2.

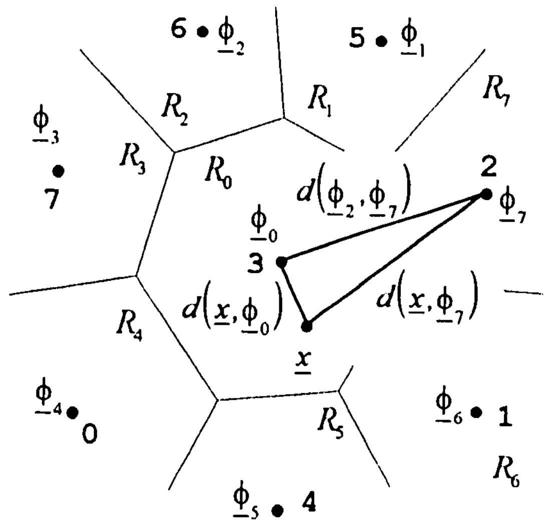
שיבוץ האינדקסים ה"טבעי" באלגוריתם זה הוא שאינדקס $i$ ישובץ לתא החלוקה $R_i^{(L)}$ ולוקטור הייצוג $\underline{\phi}_i^{(L)}$ עבור $i = 0,1,\ldots,2^l - 1$

### 2.1.2 תכן קוונטייזר וקטורי לפי Kohonen

גישה אחרת לתכן קוונטייזר וקטורי היא לפי Kohonoen [Ko90], וידועה בשם Self-Organizing Map. תהליך התכן הוא איטרטיבי ומבוסס גם כן על סדרת אימון.
בשלב ראשון בונים, על סמך ידע מוקדם או בשיטה אקראית, קוונטייזר ראשוני עם וקטורי הייצוג $i = 0,1,\ldots,N-1$, $\underline{\phi}_i^{(0)}$. בהמשך משנים את הקוונטייזר על סמך האיברים בסדרת האימון. נסמן את וקטורי הייצוג באיטרציה ה-$k$ ע"י $i = 0,1,\ldots,N-1$ $\underline{\phi}_i^{(k)}$.
עבור כל וקטור $\underline{x}$ מסדרת האימון משווים את העיוות $d\left(\underline{x},\underline{\phi}_i^{(k)}\right)$ עבור כל וקטור ייצוג. הוקטור עם העיוות הקטן ביותר מוכרז "מנצח". וקטורי ייצוג במרחק מוגדר (סביבה כדורית) מוקטור הייצוג ה"מנצח" נקראים "שכנים של המנצח".
האלגוריתם מגדיר כללי שינוי לכל אחת משלושת הקבוצות: "מנצח", "שכנים" ו"שאר העולם". הכללים מתירים שינויים גדולים יחסית בתחילת האלגוריתם. בהמשך, השינויים המתבצעים הולכים וקטנים בהתאם לפונקצית "קפיאה" (בדומה לאלגוריתם ה-Simulated Annealing שיתואר בהמשך). עבור פונקציות קפיאה איטיות, מתכנס האלגוריתם לקוונטייזר המקיים את התנאים ההכרחיים לאופטימליות.

## 2.2 קוונטיזציה וקטורית תחת שגיאות ערוץ

עד כה הנחנו כי הערוץ דרכו משודרים האינדקסים הינו חסר-שגיאות. כאשר מתרחש מאורע של שגיאה בערוץ, אינדקס שגוי נקלט במקלט ווקטור ייצוג שגוי מסופק ליעד. מאורע שגיאה מתואר בציור 2.2.




ציור 2.2 - מאורע של שגיאת-ערוץ בקוונטייזר וקטורי

Fig. 2.2 - Channel-Error event in a Vector Quantizer

הוקטור $\underline{x} \in R_0$ ולכן ישודר בערוץ האינדקס 3 המתאים לוקטור הייצוג $\underline{\phi}_0$. בעקבות השגיאה נקלט האינדקס 2, והיעד מוזן בוקטור הייצוג $\underline{\phi}_7$. בציור מתוארים הגדלים הבאים:

א. <u>עיוות הקוונטיזציה</u> $d(\underline{x}, \underline{\phi}_0)$ - העיוות ללא מאורע של שגיאת ערוץ.

ב. <u>עיוות כולל</u> $d(\underline{x}, \underline{\phi}_7)$ - העיוות הכולל במאורע המתואר.

ג. <u>עיוות הערוץ</u> $d(\underline{\phi}_0, \underline{\phi}_7)$ - המרחק בין הוקטור שנקלט לבין הוקטור שהיה אמור להיקלט ללא שגיאת ערוץ.

אנו מניחים בעבודה זו כי הערוץ הוא <u>חסר זיכרון</u> וניתן לתאר את הסתברויות המעבר בערוץ ע"י מטריצת הערוץ (Channel Transition matrix) - $Q$:

$$Q_{ij} = \text{Prob}\left(\text{Index } j \text{ Received} \,|\, \text{Index } i \text{ Transmitted}\right) \qquad (2.11)$$

שיבוץ האינדקסים לווקטורי הייצוג מיוצג ע"י מטריצת הפרמוטציה $\pi$ :

$$\mathcal{J}(\pi)=\begin{bmatrix}\text{Index assigned to } \underline{y}_0\\ \text{Index assigned to } \underline{y}_1\\ \text{Index assigned to } \underline{y}_2\\ \vdots\\ \text{Index assigned to } \underline{y}_{N-1}\end{bmatrix}=\begin{bmatrix}\mathrm{I}\underline{y}_0\\ \mathrm{I}\underline{y}_1\\ \mathrm{I}\underline{y}_2\\ \vdots\\ \mathrm{I}\underline{y}_{N-1}\end{bmatrix}=\pi\cdot\begin{bmatrix}0\\1\\2\\ \vdots\\ N-1\end{bmatrix} \tag{2.12}$$

נסתכל על המטריצה $\pi Q\pi^T$. הכפל משמאל ב-$\pi$ גורם לכך שבשורה ה-$i$ תופיעה השורה ה-$\mathrm{I}\underline{y}_i$ של המטריצה $Q$. הכפל מימין ב-$\pi^T$ גורם לכך שבעמודה ה-$j$ תופיעה העמודה ה-$\mathrm{I}\underline{y}_j$ של המטריצה $Q$. כלומר האיבר $\left(\pi Q\pi^T\right)_{ij}$ הוא ההסתברות של המאורע: שודר האינדקס המתאים ל-$\underline{y}_i$ ונתקבל האינדקס המתאים ל-$\underline{y}_j$.

לפיכך, העיוות הכולל בנוכחות שגיאות ערוץ הוא:

$$D_{\text{Total}}=\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\pi Q\pi^T\right)_{ij}\int_{R_i} d\left(\underline{x},\underline{\phi}_j\right)\cdot f_{\underline{x}}(\underline{x})\cdot d\underline{x} \tag{2.13}$$

בתא בו מתבצע האינטגרל $R_i$, הוקטור $\underline{x}$ אמור להיות מיוצג ע"י הוקטור המייצג $\underline{\phi}_i$. בגלל שגיאות הערוץ, יתכן כי וקטורים אחרים יפוענחו במקלט. ההסתברות לקליטת האינדקס המתאים לוקטור $\underline{\phi}_j$ בהינתן ששודר האינדקס המתאים לוקטור $\underline{\phi}_i$ היא כאמור $\left(\pi Q\pi^T\right)_{ij}$. עיוות הערוץ הוא ממוצע המרחק בין הוקטור שפוענח לבין הוקטור שהיה אמור להיות מפוענח ללא שגיאת ערוץ [To67]:

$$\begin{aligned}D_c&=\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\pi Q\pi^T\right)_{ij} d\left(\underline{\phi}_i,\underline{\phi}_j\right)\int_{R_i} f_{\underline{x}}(\underline{x})\cdot d\underline{x}=\\ &=\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\pi Q\pi^T\right)_{ij} d\left(\underline{\phi}_i,\underline{\phi}_j\right)p_i\end{aligned} \tag{2.14}$$

כאשר $p_i$ היא ההסתברות לשידור האינדקס המתאים לוקטור הייצוג $\underline{\phi}_i$.

בצורה מטריצית:

$$D_c = trace\left\{P\pi Q\pi^T D\right\} \qquad (2.15)$$

כאשר $P$ היא מטריצה אלכסונית המכילה את הסתברויות המקור: $P = diag\{p_0, p_1, \ldots, p_{N-1}\}$. המטריצה $D$ היא מטריצה סימטרית המכילה את המרחקים בין וקטורי הייצוג: $D_{ij} = d\left(\underline{\phi}_i, \underline{\phi}_j\right)$.

עבור ערוץ בינארי סימטרי מועבר המידע הספרתי ע"י $L = log_2 N$ סיביות. כאשר הסתברות שגיאה בערוץ $q$ שואפת לאפס, ניתנת ב-[Fa90] נוסחה לחישוב ממוצע עיוות הערוץ <u>מעל כל השיבוצים האפשריים</u> - $\overline{D}_c$.

$$\lim_{q\to 0}\frac{\overline{D}_c}{q} = \frac{L}{K(N-1)}\sum_{i=0}^{N-1} p_i \sum_{j=0}^{N-1} d\left(\underline{\phi}_i, \underline{\phi}_j\right) \qquad (2.16)$$

[To67] מראים כי עבור מידת מרחק ריבועית העיוות הכולל הוא (ראה (2.13)):

$$\begin{aligned}
D_{\text{Total}} &= \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\pi Q\pi^T\right)_{ij}\int_{R_i}\left\|\underline{x}-\underline{\phi}_j\right\|^2 \cdot f_{\underline{x}}(\underline{x})\cdot d\underline{x} = \\
&= \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\pi Q\pi^T\right)_{ij}\int_{R_i}\left\|\underline{x}-\underline{\phi}_i+\underline{\phi}_i-\underline{\phi}_j\right\|^2 \cdot f_{x}(\underline{x})\cdot d\underline{x} = \\
&= \underbrace{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\pi Q\pi^T\right)_{ij}\int_{R_i}\left\|\underline{x}-\underline{\phi}_i\right\|^2 \cdot f_{\underline{x}}(\underline{x})\cdot d\underline{x}}_{D_Q} + \\
&\quad + \underbrace{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\pi Q\pi^T\right)_{ij}\left\|\underline{\phi}_i-\underline{\phi}_j\right\|^2 \int_{R_i} f_{\underline{x}}(\underline{x})\cdot d\underline{x}}_{D_C} + \\
&\quad + 2\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\pi Q\pi^T\right)_{ij}\left(\underline{\phi}_i-\underline{\phi}_j\right)^T \int_{R_i}\left(\underline{x}-\underline{\phi}_i\right)\cdot f_{\underline{x}}(\underline{x})\cdot d\underline{x}
\end{aligned} \qquad (2.17)$$

האיבר הראשון בסכום הוא עיוות הקוונטיזציה, $D_Q$ , המחובר השני הוא עיוות הערוץ, $D_c$, והמחובר השלישי כולל איברים מצטלבים.

עבור קוונטייזר המקיים את תנאי הצנטרואיד, האינטגרל במחובר השלישי מתאפס לפי (2.10) ואז:

$$D_{\text{Total}} = D_Q + D_c \tag{2.18}$$

פיתוח דומה מופיע גם ב-[Ry76]. ראוי להוסיף כי עבור קוונטייזרים עם מספר הולך וגדל של וקטורי ייצוג ($N \to \infty$) המחובר השלישי שואף לאפס גם בקוונטייזרים שאינם מקיימים את תנאי הצנטרואיד [Ze94].

בספרות קיימות שלוש גישות להקטנת העיוות בתנאי שגיאות בערוץ:

1. **הקטנת עיוות הערוץ ע"י שיבוץ אינדקסים** - עיוות הערוץ תלוי בשיבוץ האינדקסים לוקטורי הייצוג (כלומר במטריצת הפרמוטציה $\pi$). כאמור, קיימים $N!$ שיבוצים אפשריים. עבור קוונטייזר של 4 סיביות בלבד (16 וקטורי-ייצוג) ישנם $16! \approx 2\cdot 10^{13}$ שיבוצים אפשריים. עבור קוונטייזר בן 5 סיביות ישנם $32! \approx 3\cdot 10^{35}$ שיבוצים שונים.
   בגלל מספרם הרב, לא ניתן לבחון את כל השיבוצים, ולכן הוצעו שיטות שונות תת-אופטימליות למציאת שיבוץ טוב יותר. חלק מן השיטות מתוארות בסעיף 2.3. ניתן לציין במיוחד את [De87], [Ze87], [Fa90] ואת [Kn92].

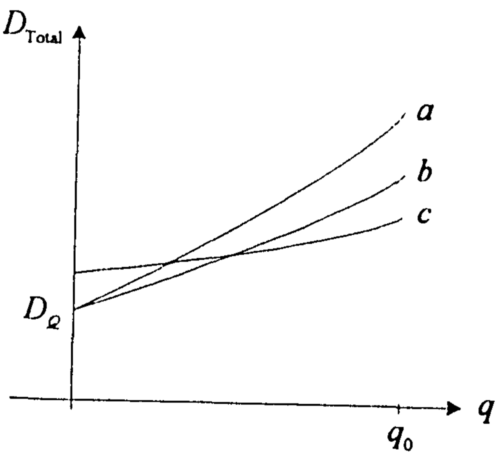
   הגישה הכללית לשיפור השיבוץ נקראת בספרות בשם Robust Vector Quantization.

2. **שינוי תאי החלוקה ווקטורי הייצוג** - ניתן להמיר את תנאי השכן הקרוב ותנאי הצנטרואיד לתנאים הלוקחים בחשבון את תנאי השגיאה בערוץ. ניתן לציין במיוחד תרומות של [Ku69], [Ze88], [Fa90] ואת [Fa91].
   גישה זו נקראת בספרות Channel Optimized Vector Quantization. גישה זו תוצג בסעיף 2.4.

3. תכן משולב של השיבוץ ותאי החלוקה - שילוב של שתי הגישות הנ"ל. נציין את המקורות [Fa87] ואת [Ze88]. גישה זו תוצג בסעיף 2.5.

התנהגות טיפוסית של העיוות הכולל מתוארת בצורה סכימטית ב-[Kn92]. בסכימה (ציור 2.3) מופיעה ההתנהגות עבור:

1. קוונטייזר "מקורי" (a)
2. קוונטייזר עם שיבוץ טוב יותר (b)
3. קוונטייזר שעבר שינוי איזור חלוקה עבור הסתברות שגיאה נתונה בערוץ שערכה $q_0$ (c)




ציור 2.3 - התנהגות טיפוסית של העיוות הכולל עבור (a) - קוונטייזר וקטורי מקורי, (b) - קוונטייזר עם שיבוץ טוב יותר ו-(c) קוונטייזר עם תאי חלוקה המתאימים להסתברות שגיאה בערוץ שערכה $q_0$

Fig. 2.3 - Typical Overall distortion characteristics for (a) Original Vector Quantizer, (b) VQ with better Index Assignment, and (c) Channel optimized VQ, designed for Channel Bit Error Rate $q_0$

כאשר קצב השגיאות בערוץ שואף לאפס אין השפעה של השיבוץ על העיוות הכולל והוא שואף לעיוות הקוונטיזציה. שיבוץ טוב יותר יביא לעיוות קטן יותר בתנאי שגיאות ערוץ. שינוי איזורי ההחלטה ווקטורי הייצוג עבור $q_0$ תרם לעיוות קטן יותר בנקודה זו. מצד שני קוונטייזר זה שונה מן התכן המקורי, האופטימלי ללא שגיאות ערוץ, ולכן יתפקד פחות טוב מן התכן המקורי כאשר אין שגיאות ערוץ.

## 2.3 הקטנת עיוות הערוץ ע"י שיבוץ אינדקסים

ביטוי כללי לעיוות הערוץ נתקבל ב- (2.15): $D_c = trace\{P\pi Q\pi^T D\}$. עבור קוונטייזר, מקור וערוץ נתונים, המתכנן מעונין לבחור בשיבוץ שיביא למינימום את עיוות הערוץ מעל כל השיבוצים, $\pi$, האפשריים. בעיית שיבוץ זו ידועה בתורת חקר-הביצועים כבעיית QA - Quadratic Assignment [Gi62]. בעיית QA ידועה כבעלת סיבוכיות NP-complete [Bu79]. עבור קוונטייזר בן $N$ וקטורי ייצוג יש צורך לבדוק $N!$ שיבוצים אפשריים.

חישוב הממוצע של עיוות הערוץ מעל כל השיבוצים האפשריים, תחת ערוץ בינארי סימטרי, מופיע ב-[Ze94]. בנוסף, קיימת במאמר הוכחה כי הממוצע **איננו שואף לאפס** גם עבור מספר רב של וקטורי ייצוג ($N \to \infty$). לשינוי במספר וקטורי הייצוג ישנן שתי השלכות:

א. עיוות הקוונטיזציה קטן עם תוספת וקטורי ייצוג.

ב. תוספת וקטורי ייצוג דורשת הגדלת האורך של מילות הקוד בערוץ ותוספת זו מגדילה את ה"פגיעות" שלהן לשגיאות. מכאן שעיוות הערוץ גדל עם תוספת של וקטורי ייצוג.

עבור קצב שגיאות נתון בערוץ, $q$, קיים מינימום לעיוות הכולל עבור $N_{min}(q)$ וקטורי ייצוג. בערוץ חסר שגיאות $q=0$, נותר רק עיוות הקוונטיזציה, ולכן $N_{min}(0) \to \infty$. עם עליית קצב השגיאות בערוץ, גדל עיוות הערוץ ו- $N_{min}(q)$ הולך וקטן.

בנוסף לממוצע מעל כל השיבוצים האפשריים, מתאר [Ze94] את הווריאנס של עיוות הערוץ מעל כל השיבוצים האפשריים ומראה כי הוא שואף לאפס כאשר $N \to \infty$. עבור הגרלה אקראית של שיבוצים נקבל לכן ריכוז חזק של ביצועים סביב הממוצע. בגלל המספר הרב של שיבוצים אפשריים, בחירה אקראית של שיבוצים תתרכז סביב הממוצע ולא תביא למציאת שיבוץ טוב בזמן סביר.

תוצאות סימולציה המתוארות ב-[Kn95] מציגות בעיה נוספת. המחבר מציג היסטוגרמה של מספר השיבוצים שנתקבלו כפונקציה של עיוות ערוץ. מתברר כי העיוות הממוצע קרוב יותר לעיוות המתקבל כתוצאה מן השיבוץ הגרוע ביותר מאשר לעיוות המתקבל מן השיבוץ הטוב ביותר. קיים "זנב" בעל שכיחות נמוכה לכיוון השיבוץ הטוב ביותר, אשר מחליש את המוטיבציה לשימוש בשיבוץ אקראי.

בגלל המספר הרב של שיבוצים אפשריים, הוצעו בספרות שיטות תת-אופטימליות לשיבוץ אינדקסים. שיטות אלו מבוססות על שיפור השיבוץ ע"י החלפת זוגות. הוצע גם אלגוריתם המנסה להביא לשיבוץ גלובלי - Simulated Annealing. סקירה קצרה של גישות אלה נתונה להלן.

### 2.3.1 שיבוץ אינדקסים ע"י אלגוריתם החלפת זוגות

עבור ערוץ בינארי סימטרי מאורע השגיאה הדומיננטי הוא מאורע של שגיאה בסיבית אחת. נקרא לזוג אינדקסים המרוחקים ביניהם בסיבית אחת (מרחק Hamming שווה לאחד) - אינדקסים שכנים. לכל אינדקס יש $L$ שכנים, כמספר הסיביות בקוונטייזר.

שיטה הכוללת שיבוץ ראשוני אד-הוק ושיפור השיבוץ ע"י חילופי אינדקסים מתוארת ב-[De87] .

אלגוריתמים דומים מאד במהותם הוצגו ע"י [Ze87], [Ze90] ו-[Ch93]. האינדקס הראשון (0) משובץ לוקטור ייצוג בעל ההסתברות הגבוהה ביותר. בשלב שני משובצים האינדקסים השכנים לאינדקס הראשון לוקטורי ייצוג הקרובים לוקטור הראשון. שאר האינדקסים משובצים לוקטורי-ייצוג בני הסתברות הולכת ויורדת. כל אינדקס משובץ לוקטור הפנוי הקרוב ביותר לוקטורים שכבר שובצו עם אינדקסים שכנים. בשלב שני נעשה ניסיון לשפר את השיבוץ ע"י החלפת אינדקסים. אלגוריתם השיפור עובר על זוגות אינדקסים ומנסה להחליף ביניהם. מתקבל שיבוץ חדש אשר שונה מן השיבוץ הקודם בשני אינדקסים. אם השיבוץ החדש מביא לעיוות ערוץ נמוך יותר, האלגוריתם מאמץ אותו. החלפת האינדקסים עבור זוגות ממשיכה עד אשר לא ניתן יותר לשפר את עיוות הערוץ בצורה זו. האלגוריתם מבטיח התכנסות לאופטימום מקומי בלבד.

ב-[Ze90] מוצגות גם תוצאות מספריות עבור מקורות מסוג גאוס-מרקוב. עבור מקור גאוס-מרקוב עם מקדם מתאם של 0.5 נבנה קוונטייזר בן 8 סיביות עם וקטורי-ייצוג בני מימד 4 (2 סיביות לדגימה). עבור הסתברות שגיאה של 1% בערוץ בינארי סימטרי נתגלה הבדל של כ-5dB בין שיבוצים גרועים לשיבוצים טובים (ראה גם ציור 6.5 בעבודה זו).

מערכת לקידוד אות דיבור מסוג Vector Adaptive Predictive Coding מתוארת ב-[Ch87]. המערכת משתמשת בקידוד LPC - Linear Predictive Coding לאות הדיבור. אות העירור (Excitation/Residual) מיוצג ע"י קוונטייזר וקטורי. שיבוץ האינדקסים לוקטורי היצוג נעשה בסוף תהליך התכן ע"י אלגוריתם דומה של החלפת אינדקסים.

### 2.3.2 שיבוץ אינדקסים ע"י אלגוריתם Simulated Annealing

שיטת השיבוץ ע"י החלפת זוגות של אינדקסים נותנת עיוות עורץ היורד מונוטונית עם התקדמות האלגוריתם, אך מבטיחה התכנסות למינימום מקומי בלבד. ניסיון למציאת מינימום גלובלי לעיוות הערוץ מעל שיבוצי האינדקסים נעשה ב-[Fa90] תוך שימוש באלגוריתם של - SA Simulated Annealing. נתאר בהמשך בסעיף 2.5 מקור אחר [Cu94] העושה שימוש דומה בטכניקת SA לשיבוץ אינדקים. אלגוריתם ה-SA מנסה למנוע את בעיית ההתכנסות למינימום מקומי בכך שניתן (בהסתברות הולכת ויורדת) לצאת ממינימום כזה. המוטיבציה לשימוש באלגוריתם ה-SA באה מתורת הגבישים. החומר ממנו עשוי הגביש מומס ומקורר לאט. הקרור האיטי מאפשר למולקולות החומר להסתדר במבנה גבישי ולהגיע לאנרגיה פוטנציאלית מינימלית. עבור בעיות קומבינטוריות קיימת פונקצית מחיר (שקולה לאנרגיה פוטנציאלית) במשתנים רבים. המשתנים קשורים ביניהם באילוצים (שקולים לקשרים כימיים). עבור האלגוריתם מגדירים גורם "טמפרטורה". גורם זה מאפשר בשלבים הראשונים של האלגוריתם לשנות במידה רבה את מצב המערכת. בהמשך מונמכת השפעת גורם ה"טמפרטורה", המערכת "מתקררת" ומתייצבת.

במאמר הוצע האלגוריתם הבא:

**<u>אלגוריתם שיבוץ אינדקסים מבוסס Simulated Annealing עפ"י [Fa90]</u>**

1. התחל בגורם טמפרטורה $T \leftarrow T_0$, ובשיבוץ התחלתי $\pi \leftarrow \pi_0$.
2. החלף באקראי זוג אינדקסים בשיבוץ וקבל שיבוץ חדש $\pi'$.

   סמן את השינוי בעיוות הערוץ ע"י $\Delta D_c = D_c(\pi') - D_c(\pi)$.

   אם השיבוץ החדש מביא לעיוות ערוץ קטן יותר יש לאמץ אותו $\pi \leftarrow \pi'$.

   אחרת יש לאמץ אותו בהסתברות $exp(-\Delta D_c/T)$. (ההסתברות קטנה עם ירידת $T$)
3. אם בשלב 2 נעשו יותר ממספר מוגדר של שינויים עבור לשלב 4. אחרת חזור על 2.
4. הנמך את הגורם $T$. אם הגעת לטמפרטורה מוגדרת ("קפיאה") הפסק. אחרת חזור לשלב 2.

אלגוריתם ה-SA הינו יוריסטי וקיימים בו גורמים רבים המוגדרים בהתאם לדמיונו של המתכנן. כותב המאמר עמל קשה לקבלת תוצאות טובות ע"י אלגוריתם "קירור" שנבחן אמפירית. חיסרון נוסף של האלגוריתם הוא בכך ש-"קירור" האלגוריתם חייב להיות איטי להשגת תוצאות טובות, ולכן קצב ההתכנסות שלו איטי.

המאמר מתאר ניסויים רבים בקוונטייזרים וקטוריים במימדים שונים, כולם בקצב של 1 סיבית לדגימה, שתוכננו למקור גאוס-מרקוב עם מקדמי מתאם 0.0 (חסר זיכרון) ו-0.9. לדוגמה, עבור מקדם מתאם 0.9, קוונטייזר בן 8 סיביות נתן רווח של כ-5dB מן הממוצע של עיוות הערוץ על פני כל השיבוצים האפשריים.

במאמר קיימת הבחנה מעניינת הנוגעת לאלגוריתם ה-LBG במבנה עץ שתואר בסעיף 2.1. מחבר המאמר מצא, אמפירית, כי שיבוץ האינדקסים ה-"טבעי" לווקטורי הייצוג נתן עיוות ערוץ קטן מן הממוצע של עיוות הערוץ על פני כל השיבוצים האפשריים.

## <u>2.4 שינוי תאי החלוקה ווקטורי הייצוג</u>

תכן הקוונטייזר ללא התחשבות בשגיאות ערוץ נעשה, כזכור, בעזרת שני תנאים הכרחיים למינימום: תנאי השכן הקרוב ותנאי הצנטרואיד. יישום לסירוגין של שני התנאים הוא הבסיס לאלגוריתמים לתכן הקוונטייזר. בתנאי שגיאות ערוץ ניתן לפתח תנאים דומים המשקללים את העיוות לפי הסתברויות המעבר לכל הוקטורים המייצגים. הפיתוחים בסעיף זה מניחים כי גם למשדר וגם למקלט מידע על הערוץ.

### <u>2.4.1 קוונטייזר סקלרי</u>

פיתוח התנאים ההכרחיים עבור קוונטייזר סקלרי הפועל תחת שגיאות ערוץ נעשה ב-[Ku69] עבור מידת שגיאה ריבועית. עבור קוונטייזר סקלרי, תנאי השכן הקרוב המשוקלל, בהתחשב בשגיאות

הערוץ, מתבטא ברמות-החלטה $i=0,1,\ldots,N$ , $\tau_i$, כאשר $\tau_0=-\infty$ ו-$\tau_N=+\infty$, והאחרים נתונים ע"י:

$$\tau_i=\frac{1}{2}\frac{\sum_{j=0}^{N-1}\phi_j^2\left(Q_{ij}-Q_{i-1,j}\right)}{\sum_{j=0}^{N-1}\phi_j\left(Q_{ij}-Q_{i-1,j}\right)},\quad i=1,2,\ldots,N-1 \tag{2.19}$$

עבור ערוץ מושלם $Q=I$, מצטמצם הביטוי לכדי $\tau_i=\frac{1}{2}\left(\phi_i+\phi_{i+1}\right)$.

רמות הייצוג, $i=0,1,\ldots,N$ , $\phi_i$, מתקבלות ע"י תנאי הצנטרואיד המשוקלל:

$$\phi_i=\frac{\sum_{j=0}^{N-1}Q_{ij}\cdot\int_{\tau_j}^{\tau_{j+1}}x\cdot p(x)dx}{\sum_{j=0}^{N-1}Q_{ij}\cdot\int_{\tau_j}^{\tau_{j+1}}p(x)dx},\quad i=0,1,\ldots,N-1 \tag{2.20}$$

עבור ערוץ מושלם $Q=I$, מצטמצם הביטוי להיות (2.10):

$$\phi_i=\frac{\int_{\tau_i}^{\tau_{i+1}}x\cdot p(x)dx}{\int_{\tau_i}^{\tau_{i+1}}p(x)dx},\quad i=0,1,\ldots,N-1 \tag{2.21}$$

יש להדגיש כי יישום התנאים לסירוגין, בנוכחות שגיאות ערוץ, בדומה לשיטת התכן המקורית של קוונטייזר איננו מבטיח התכנסות האלגוריתם, אפילו לא למינימום מקומי. לחילופין, מציעים המחברים לבצע בשלב ראשון תכן של קוונטייזר עבור ערוץ ללא שגיאות. בהמשך מעלים בצעדים איטיים את הסתברות השגיאה בתנאים (2.19) ו-(2.20) עד להתייצבות. בכל צעד המבנה שנוצר מהווה תנאי התחלה לצעד התכן הבא.

באופן דומה מגדירים המחברים את בעיית התכן לקוונטייזר סקלרי אחיד עבור שגיאות ערוץ. פתרון המשוואות המתקבלות למציאת צעד הקוונטיזציה הינו נומרי.

### 2.4.2 קוונטייזר וקטורי

עבור קוונטייזר וקטורי הוצעו בספרות ([Ze88], [Fa90] ו-[Fa91]) תנאים דומים לתכן: תנאי השכן הקרוב המשוקלל ותנאי הצנטרואיד המשוקלל.

### תנאי השכן הקרוב המשוקלל (Weighted Nearest Neighbor Condition)

בהינתן ספר קוד $C$, וקטור כניסה $\underline{x}$ ישוחזר ע"י וקטור הייצוג הקרוב אליו ביותר בשקלול שגיאות הערוץ:

$$Q(\underline{x}) = \underline{\phi}_i \quad \text{if} \quad \sum_{k=0}^{N-1} Q_{ik} d(\underline{x}, \underline{\phi}_k) \le \sum_{k=0}^{N-1} Q_{jk} d(\underline{x}, \underline{\phi}_k) \quad \text{for } j = 0,1,\ldots,N-1 \tag{2.22}$$

תנאי זה קובע את תאי החלוקה $R_i$ , $i = 0,1,\ldots,N-1$. הבדיקה לאיזה תא שייך וקטור כניסה $\underline{x}$ נעשית ע"י חישוב המרחק המשוקלל ממנו לכל וקטורי הייצוג ומציאת הקרוב אליו ביותר, במובן של מידת המרחק המשוקללת.

### תנאי הצנטרואיד המשוקלל (Weighted Centroid Condition)

בהינתן חלוקת המרחב לתאי חלוקה, וקטור ייצוג $\underline{\phi}_i$ המייצג תא חלוקה $R_i$ הוא הוקטור המביא למינימום העיוות בתא זה:

$$\begin{aligned} &Q(\underline{x}) = \underline{\phi}_j \quad \text{for all } \underline{x} \in R_j \\ &\text{if} \sum_{i=0}^{N-1} Q_{ij} \cdot \mathrm{E}\left\{ d(\underline{x}, \underline{\phi}_i) \,\middle|\, \underline{x} \in R_i \right\} \le \sum_{i=0}^{N-1} Q_{ij} \cdot \mathrm{E}\left\{ d(\underline{x}, \underline{z}) \,\middle|\, \underline{x} \in R_i \right\} \quad \text{for all } \underline{z} \end{aligned} \tag{2.23}$$

התנאים ההכרחיים (2.22) ו-(2.23) אינם מבטיחים כי הקוונטייזר הוקטורי יהיה רגולרי. בנוסף לכך, יישום שני התנאים לסירוגין איננו מבטיח התכנסות. גם במקרה זה הפתרון המעשי הוא תכן קוונטייזר ללא התחשבות בשגיאות ערוץ בשלב ראשון. בהמשך מעלים בהדרגה את קצב השגיאות בערוץ עד לקצב הרצוי, כפי שתואר קודם. הסתכלות על קוונטייזר וקטורי שתוכנן בהתחשב בשגיאות ערוץ מראה כי תאי החלוקה ווקטורי הייצוג "מתכווצים" לכיוון "מרכז ההסתברות" של האות.

## 2.4.3 תכן קוונטייזר וקטורי לפי Kohonen

גישה אחרת לתכן קוונטייזר וקטורי תחת שגיאות ערוץ מופיעה ב-[Kn92]. הגישה מבוססת על השיטה לתכן קוונטייזר וקטורי לפי Kohonoen, שתוארה בסעיף 2.1.2.

השינוי באלגוריתם הוא החלפת העיוות $d\left(\underline{x}, \underline{\phi}_i^{(k)}\right)$ בעיוות המשוקלל $\sum_{j=0}^{N-1} Q_{ij}\, d\left(\underline{x}, \underline{\phi}_i^{(k)}\right)$. שימוש מושכל בפונקציה הקפיאה מסייע בהתכנסות אלגוריתם.

## 2.5 תכן משולב של מבנה הקוונטייזר ושיבוץ האינדקסים

ניתן לשלב את שתי הגישות לשיפור ביצועי הקוונטייזר הוקטורי תחת שגיאות ערוץ [Ze88]. בשלב ראשון יוצרים שיבוץ אקראי של אינדקסים. עבור שיבוץ זה מתכננים קוונטייזר המשקלל את השפעת שגיאות הערוץ. כעת ניתן לשפר לסירוגין את השיבוץ ואת מבנה הקוונטייזר עד להתכנסות.

גישה ל״שחזור רד״ (SDVQ - Soft Decision Vector Quantization) מתוארת ב-[Cu94]. בגישה שתיארנו לשינוי תאי החלוקה ווקטורי הייצוג, ה-COVQ , תמיד משוחזר וקטור ייצוג מתוך ספרית וקטורי הייצוג הקיימת. גישת ה״שחזור הרד״ מציעה להתנתק מאילוץ זה. כמקודם, המקודד מחלק את קבוצת כל הוקטורים האפשריים לתאי חלוקה $R_i$. וקטור כניסה $\underline{x}$ השייך לתא חלוקה $R_i$ מיוצג בערוץ ע״י סימן ערוץ $\underline{s}_i$. בכניסת המפענח מתקבל משתנה אקראי $\underline{y}$. המפענח מבצע שיערוך של וקטור הכניסה ע״י:

$$\hat{\underline{x}}(\underline{y}) = E[\underline{x}|\underline{y}] = \int_{\Re^K} \underline{x}\, p(\underline{x}|\underline{y})\, d\underline{x} =$$

$$= \frac{1}{p(\underline{y})}\int_{\Re^K} \underline{x}\, p(\underline{x})\, p(\underline{y}|\underline{x})\, d\underline{x} = \frac{1}{p(\underline{y})}\sum_{i=0}^{N-1}\int_{R_i} \underline{x}\, p(\underline{x})\, p(\underline{y}|\underline{s}_i)\, d\underline{x} =$$

$$= \frac{1}{p(\underline{y})}\sum_{i=0}^{N-1} p(\underline{y}|\underline{s}_i)\, p_i\, \underline{\phi}_i = \frac{\sum_{i=0}^{N-1} p(\underline{y}|\underline{s}_i)\, p_i\, \underline{\phi}_i}{\sum_{i=0}^{N-1} p(\underline{y}|\underline{s}_i)\, p_i} \qquad (2.24)$$

כאשר $p_i$ היא ההסתברות להופעת אות הערוץ $\underline{s}_i$ ו-$\underline{\phi}_i$ הוא הצנטרואיד של התא $R_i$ (שאיננו וקטור שחזור במקרה זה). המפענח משחזר שיקלול של הצנטרואידים לפי ההסתברות להופעת הוקטור $\underline{y}$ עבור כל תא חלוקה. המחברים מציגים תנאים אופטימליים לתאי החלוקה עבור ערוץ גאוסי אדיטיבי: $\underline{y} = \underline{s} + \underline{n}$, כאשר $\underline{s}$ הוא אחד מאותות הערוץ האפשריים ו-$\underline{n}$ הוא רכיב הרעש. חלוקת אותות הערוץ לתאים המתאימים נעשית ע״י Simulated Annealing.

בהמשך המאמר מציעים המחברים הרחבה לשיטת ה-SDVQ, המנצלת תלות בין וקטורי מקור עוקבים. המשערך לוקטור הכניסה בהינתן מוצא הערוץ מתבסס בזמן $t$ גם על מוצא הערוץ בזמנים שכנים: $\hat{\underline{x}}(\underline{y}_{t-T_1},\ldots,\underline{y}_{t-1},\underline{y}_t,\underline{y}_{t+1},\ldots,\underline{y}_{t+T_2})$.

בדוגמאות מספריות המוצגים במאמר מציג ה-SDVQ ביצועים עדיפים על ה-COVQ.

$(N \leq M)$ רמות ייצוג. עבור מקודד נתון, נעשה שימוש במפענח המשערך את רמת הכניסה בדומה ל-(2.24). עבור המקודד מתקבל תנאי השכן הקרוב המשוקלל ותנאי הצנטרואיד המשוקלל. המחברים מתארים את הסיבה לבעיית אי-ההתכנסות בתכן קוונטייזרים סקלריים המתחשבים בשגיאות הערוץ. הבעייה נובעת מכך שלא מובטח בזמן התכן כי רמות ההחלטה לא יתחלפו בסדר. כלומר גם אם בתנאי ההתחלה רמות ההחלטה מסודרות בסדר עולה $\tau_i \leq \tau_{i+1}$ , הרי שייתכן שבשלב מסוים התנאים המשקללים יהפכו את הסדר. מסיבה זו מוסיפים המחברים את תנאי הסדר באלגוריתם התכנון. תנאי זה גורם לרמות החלטה מסוימות להיעלם $(\tau_i = \tau_{i+1})$ ולהקטין בכך את מספר רמות הייצוג $N$. תרחיש כזה מתקיים כאשר עיוות הערוץ גדול מספיק על מנת להצדיק הורדת רזולוציה בקידוד המקור והוספת גורם תיקון שגיאות בערוץ. במאמר מוצגת דוגמה עבור מקור גאוסי ו-$M = 8$ סימני ערוץ (ערוץ בינארי סימטרי, 3 סיביות לדגימה) אפשריים. עבור הסתברויות שגיאה נמוכות ישנם $N = 8$ תאי חלוקה. רמות הייצוג וההחלטה "מתכווצות" לכיוון האפס עד להסתברות שגיאה בערוץ $q \approx 0.03$. בהסתברויות שגיאה גבוהות יותר מתאחדות רמות החלטה ויוצרות $N = 4$ תאי חלוקה בלבד.

ההתנהגות האסימפטוטית של העיוות הכולל במקודד משולב עם $N$ וקטורי ייצוג ו-$M$ סימני ערוץ, עבור מספר וקטורי ייצוג השואף לאינסוף מתוארת ב-[Ze94]. מידת העיוות הנבדקת היא המרחק האוקלידי בחזקת $r$, $\|X - Q(X)\|^r$. בהנחות המוצגות במאמר, יורד עיוות הקוונטיזציה לאפס לפחות לפי $2^{-rN/K}$ כאשר $K$, כזכור, הוא מימד הוקטורים עליהם מתבצע הקידוד. עיוות הערוץ יורד לאפס לפחות לפי $2^{-NE_{max}\left(\frac{log_2 N}{log_2 M}\right)}$, כאשר $E_{max}(\cdot)$ היא פונקצית האמינות המוגדרת עבור קידוד ערוץ [Vi79]. פונקציה זו היא חיובית עבור קצבים הנמוכים מקיבול הערוץ $C$ $\frac{log_2 N}{log_2 M} < C$.

המסקנה היא שעבור $N \to \infty$ ניתן לבחור $M > N$ אשר יבטיח עיוות כולל השואף לאפס.

מן הראוי לציין כי איננו עוסקים בעבודת מחקר זו בתכן משולב מקור-ערוץ מן הסוג המתואר בסעיף זה.

# פרק 3 - חסמים על עיוות הערוץ מעל כל המקורות האפשריים

## 3.1 מבוא

בפרק המבוא תיארנו את השפעת שגיאות הערוץ על ביצועי קוונטייזרים וקטוריים תחת ערוצים חסרי זיכרון. הגדרנו את עיוות הערוץ וראינו כי בתנאים מסוימים העיוות הכולל הוא סכום של העיוותים שהוגדרו כעיוות הערוץ וכעיוות הקוונטיזציה. בפרק זה נתאר שיטות לחסימה ולהערכת-ביצועים של מערכות נתונות מעל כל המקורות האפשריים. החסמים יתקבלו תוך שימוש בטכניקות של תכנות ליניארי. עבור ערוץ בינארי סימטרי (עם או בלי קוד לתיקון שגיאות) נבחן בהמשך גם את ביצועי המערכות עבור הסתברויות שגיאה נמוכות בערוץ.

כן נוכיח מספר תכונות מעניינות של המקרה המיוחד: קוונטייזר סקלרי אחיד, קוד בינארי טבעי (NBC - Natural Binary Code) וערוץ בינארי סימטרי (BSC - Binary Symmetric Channel). נראה כי עיוות הערוץ במקרה זה כמעט ואיננו תלוי בהסתברויות המקור ובמספר הסיביות. נראה כי עבור קוונטייזרים בני 4 סיביות ומעלה, השיבוץ הבינארי הטבעי טוב מן הממוצע על פני כל השיבוצים, עבור כל מקור אפשרי.

## 3.2 חסמי ביצועים מעל כל המקורות האפשריים

ביטוי כללי לעיוות הערוץ ניתן ב-(2.15) ע"י:

$$D_c = trace\left\{P\pi Q_N \pi^T D\right\} = \sum_{i=0}^{N-1} P_{ii} \cdot \left(\pi Q_N \pi^T D\right)_{ii} \tag{3.1}$$

כאשר:

א. מטריצת הערוץ $Q$ מתארת את ההסתברויות המעבר עבור אינדקסים המשודרים בערוץ.

$$Q_{ij} = \text{Prob}\left\{ j \text{ received} \middle| i \text{ transmitted} \right\} \quad i,j = 0,1,\ldots,N-1$$

ב. מטריצת הפרמוטציה $\pi$ מתארת את שיבוץ מילות הקוד. האיבר $\left(\pi Q_N \pi^T\right)_{ij}$ הוא הסתברות קבלת האינדקס המתאים לוקטור הייצוג $\underline{\phi}_j$ בהינתן ששודר האינדקס המתאים לוקטור הייצוג $\underline{\phi}_i$.

ג. מטריצת המרחקים $D$ מתארת את המרחקים בין וקטורי הייצוג $\underline{\phi}_i$ -

$$D_{ij} = d\left(\underline{\phi}_i, \underline{\phi}_j\right) \quad i,j = 0,1,\ldots,N-1$$

ד. מטריצת ההסתברויות $P$ היא מטריצה אלכסונית המכילה את ההסתברויות להופעת וקטור בתא החלוקה $R_i$ - $i = 0,1,\ldots,N-1$ $\quad P = diag\left(\text{Prob}\left\{\underline{x} \in R_i\right\}\right)$.

עבור מערכות נתונות, ידועים הגדלים א',ב' ו-ג'. המתכנן עשוי להיות מעונין לדעת את עצמתו של עיוות הערוץ, עבור מקורות שונים. כפי שנראה ניתן באמצעים של תכנות ליניארי לקבל חסמים על עיוות הערוץ מעל כל המקורות האפשריים. בהמשך נקבל חסמים הדוקים יותר על משפחות מצומצמות יותר של מקורות.

לנוחיות הקורא מרוכזים בנספח ו' ההגדרות אשר משמשות אותנו בעבודה זו ומנוסח המשפט היסודי של התכנות הליניארי (ראה גם [Lu84]).

ניתן לחסום את עיוות הערוץ ע"י מציאת המינימום והמקסימום שלו מעל קבוצת כל המקורות האפשריים.

נציג את בעיית התכנות הליניארי הבאה:

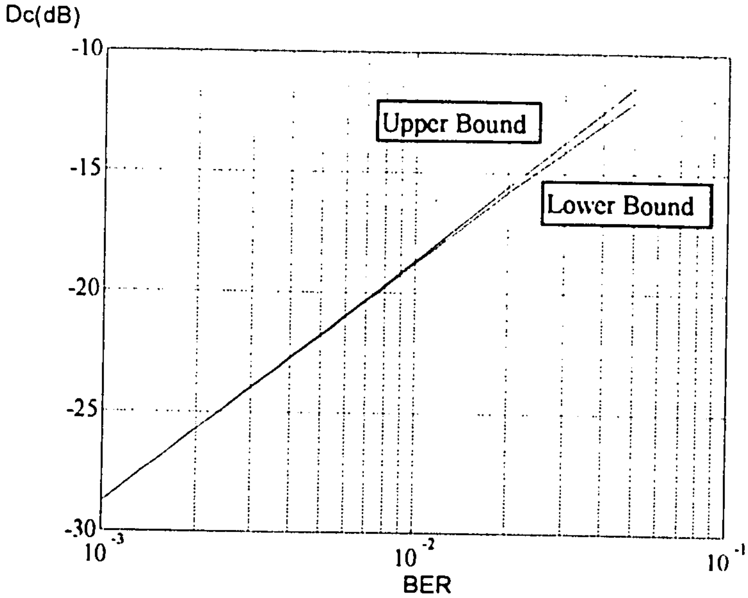
$$\begin{aligned} &\min_P / \max_P \quad \sum_{i=0}^{N-1} P_{ii} \cdot \left(\pi Q_N \pi^T D\right)_{ii} \\ &s.t. \quad \sum_{i=0}^{N-1} P_{ii} = 1 \\ &\qquad\quad P_{ii} \geq 0 \end{aligned} \tag{3.2}$$

האופטימום לבעיות המינימום והמקסימום מושגים **בפתרונות הבסיסיים** (Basic Solutions) של הבעיה (ראה נספח ו'). לבעיה זו יש $N$ משתנים ואילוץ יחיד. ישנם לכן $N$ פתרונות בסיסיים, כאשר פתרון בסיסי מתקבל ע"י איפוס של $N-1$ משתנים והשוואת המשתנה הנותר לאחד. כלומר:

$$\text{Basic solution } k \quad (k = 0,1,\ldots,N-1): \quad P_{ii} = \delta_{ik} = \begin{cases} 0 & i \neq k \\ 1 & i = k \end{cases} \tag{3.3}$$

לפתרון בעיית התכנות הליניארי עלינו לחשב את ערכי האלכסון של המטריצה $\left(\pi Q_N \pi^T D\right)$ ולאתר את האיבר הגבוה ביותר ואת האיבר הנמוך ביותר, למציאת החסם העליון והתחתון בהתאמה.

לדוגמה, עבור קוונטייזר סקלרי אחיד בן 16 רמות (4 סיביות) בתחום [-1,1], הערוץ הבינארי הסימטרי (BSC) והקוד הבינארי הטבעי (NBC) המתאים ל-$\pi = I$, חושבו החסמים, עבור הסתברויות שגיאה, $q$, שונות, והם מתוארים בציור 3.1.



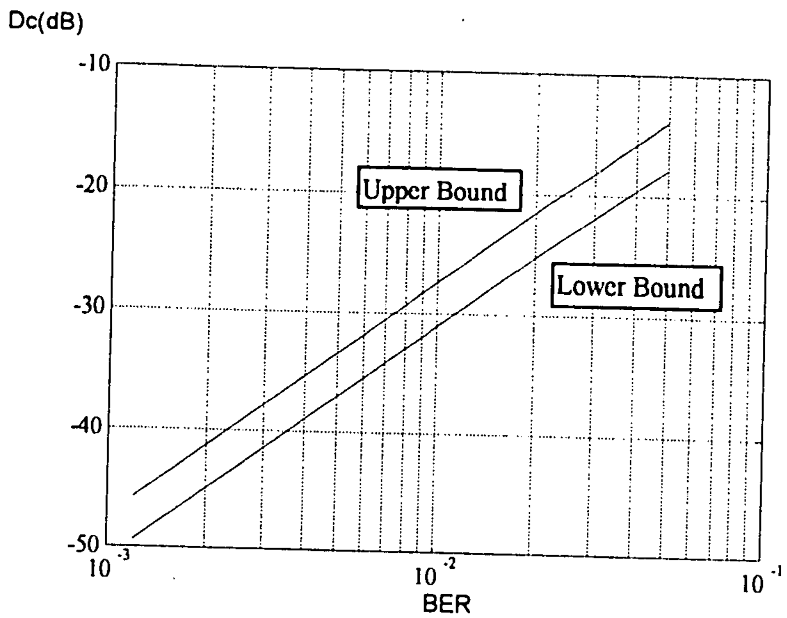

ציור 3.1 - חסמים על עיוות הערוץ מעל כל מקורות האות האפשריים
(קוונטייזר סקלרי אחיד 4 סיביות בתחום $[-1,1]$, ערוץ בינארי סימטרי, קוד בינארי טבעי)
Fig. 3.1 - Bounds on Channel-Distortion over all possible signal sources
(4 bit Uniform Scalar Quantizer covering $[-1,1]$, Binary Symmetric Channel, Natural Binary Code)

במקרה זה ניתן לראות כי החסמים (העליון והתחתון) על עיוות הערוץ הנם קרובים מאד זה לזה וכי הביצועים של הקוונטייזר האחיד עם השיבוץ הבינארי הטבעי כמעט ואינם תלויים בסוג המקור המזין אותו. הסיבה לכך היא שבקוונטייזר סקלרי אחיד עם הקוד הבינארי הטבעי לכל סיבית מתאים עיוות אוקלידי. כלומר, עבור הסיבית המשמעותית ביותר, העיוות המתאים הוא 1; עבור הסיבית הבאה בחשיבותה העיוות המתאים הוא $\frac{1}{2^2}$ וכו'. שגיאה בסיבית מסוימת (מתוך $L$ סיביות) תגרום לעיוות זהה עבור כל רמת ייצוג.

ניתן להבחין כי בסקלה לוגריתמית השיפוע של החסמים הוא כמעט קבוע ושווה ל-10dB לדקדה (הקטנת קצב השגיאות פי 10 הביאה לירידה של פי 10 בעיוות הערוץ). ניתן לכן להעריך כי החסמים יחסיים באופן ישר לשגיאה בערוץ $q$. נחזור לנקודה זו בסעיף 3.4.

ניתן להשתמש בטכניקת חסימה זו עבור <u>כל ערוץ חסר זיכרון</u>. דוגמה לערוץ כזה היא ערוץ בינארי סימטרי עם קוד לתיקון שגיאות. כאשר משדרים בערוץ את האינדקסים של וקטורי הייצוג ללא תיקון שגיאות, כל שגיאת-ערוץ גורמת להופעת וקטור ייצוג מוטעה במוצא המפענח. קוד לתיקון שגיאות משמש לגילוי ותיקון שגיאות-ערוץ שכיחות כך שלא יגרמו לוקטור ייצוג לא-מתאים להופיע במוצא המפענח. לגבי הקוונטייזר הוקטורי (והסקלרי כמקרה פרטי) מתקבל ערוץ חסר זיכרון שקול, וניתן לחשב עבורו את מטריצת הערוץ $Q$.

לדוגמה, עבור הקוונטייזר הסקלרי האחיד מן הדוגמה הקודמת ניתן לשבץ לכל רמה מלת-קוד בת 7 סיביות תוך שימוש ב-Hamming Code (7,4) [Ta86 סעיף 13.17]. במקרה זה המרחק בין מלות קוד הוא לפחות 3 סיביות, ולכן הקוד מסוגל לתקן שגיאה בסיבית אחת. החסמים על ביצועי מערכת זו מופיעים בציור 3.2:




ציור 3.2 - חסמים על עיוות הערוץ מעל כל מקורות האות האפשריים עבור קוונטייזר סקלרי אחיד 4 סיביות בתחום $[-1,1]$, ערוץ בינארי סימטרי וקוד Hamming (7,4)

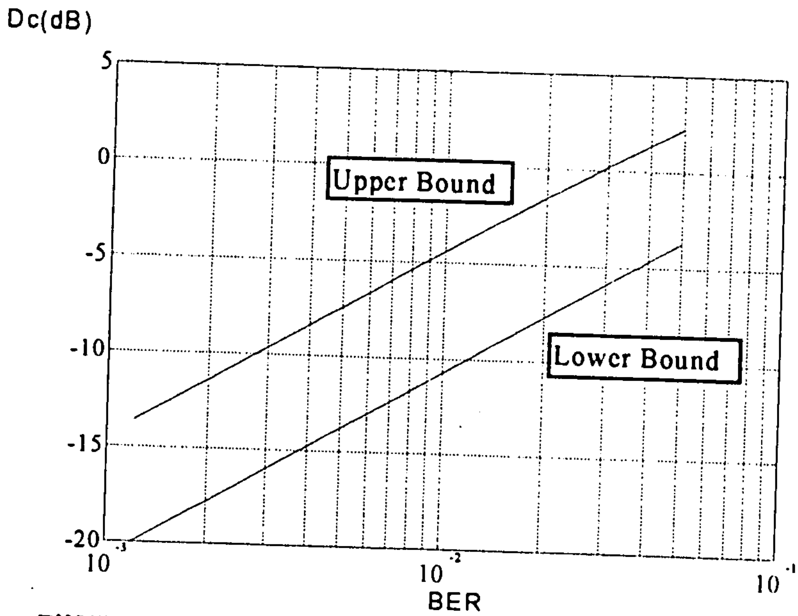
Fig. 3.2 - Bounds on Channel-Distortion over all possible signal sources for 4-bit Uniform Scalar Quantizer covering $[-1,1]$,
Binary Symmetric Channel and (7,4) Hamming Code

ניתן להבחין בירידה משמעותית (קרוב ל-12dB עבור $q = 10^{-2}$) בעיוות הערוץ כתוצאה מביטול העיוות המתקבל משגיאת סיבית בודדת (המאורע השכיח ביותר). קיים הפרש של כ-3dB בין החסם העליון והחסם התחתון עבור מקורות שונים, דבר המצביע על כך שהתכונה של התאמת עיוות לסיביות במילת הקוד נעלמה.

ניתן להבחין כי בסקלה לוגריתמית השיפוע של החסמים הוא כמעט קבוע ושווה ל-20dB לדקדה (הקטנת קצב השגיאות פי 10 הביאה לירידה של פי 100 בעיוות הערוץ). ניתן להעריך כי החסמים יחסיים לריבוע השגיאה בערוץ - $q^2$. נחזור לנקודה זו בסעיף 3.4.

נתבונן כעת בקוונטייזר וקטורי בעל 256 וקטורים מייצגים (8 סיביות), בקצב של 2 סיביות לדגימה ($K = 4$ דגימות בוקטור ייצוג). הקוונטייזר תוכנן למקור גאוס-מרקוב מסדר ראשון [Ge92 סעיף 11.4] בעל תוחלת אפס ומקדם מתאם (קורלציה) 0.5. לשם התכנון עשינו שימוש באלגוריתם ה-LBG במבנה עץ המתואר בפרק 2 [Ge92 סעיף 12.4].

בציור 3.3 מתוארים החסמים עבור קוונטייזר וקטורי זה:




ציור 3.3 - חסמים על עיוות הערוץ מעל כל מקורות האות האפשריים
(קוונטייזר וקטורי 8 סיביות, 2 סיביות לדגימה עבור מקור גאוס-מרקוב, ערוץ בינארי סימטרי)
Fig. 3.3 - Bounds on Channel-Distortion over all possible signal sources
(8 bit Vector Quantizer, 2 bit pre sample designed for Markov-Gaussian Source, Binary Symmetric Channel)

במקרה זה ניתן לראות כי עיוות הערוץ תלוי במקור המזין אותו. מטכניקת בניית החסמים ניתן להבין כי קיימים שני וקטורי ייצוג "טוב" ו-"גרוע". מקור המרוכז בוקטור הייצוג ה-"טוב" נותן עיוות ערוץ הקטן ב-8dB ממקור המרוכז בוקטור ייצוג "גרוע".

## 3.3 שיפור החסמים ע"י שימוש במומנטים

לעתים המרחק בין החסמים גדול יחסית וידיעתם איננה תורמת מספיק להערכת ביצועי המערכת. הדבר נגרם מכך שהמקורות המגשימים את החסמים אינם טיפוסיים למקור נתון. ברב המקרים מסוג זה, בהם נתקלנו, ניתן להבחין כי **החסם העליון ממומש ע"י מקור המרוכז בתאי קצה** (Overload Cell) או **באינטרוולי קצה** בקוונטייזרים סקלריים. מקור כזה איננו מייצג נאמנה מבחינת עיוות ערוץ את המקורות הטיפוסיים המוזנים לקוונטייזר.

ניתן לקבל חסמים הדוקים יותר עבור משפחה מצומצמת יותר של מקורות, המאופיינת ע"י המומנטים של המקורות:

$$\sum_{i=0}^{N-1} P_{ii} \cdot f_k\left(\underline{\phi}_i\right) \le M_k \quad k = 0,1,\ldots,K-1 \tag{3.4}$$

כאשר $f_k\left(\underline{\phi}\right)$ הן פונקציות ידועות (לרב, במקרה הסקלרי, $f_k(x) = x^k$) ו-$M_k$ הוא המומנט המתאים.

המומנטים המופיעים ב-(3.4) הנם אילוצים ליניאריים בהסתברויות המקור $P_{ii}$. אילוצים אלו מתווספים לבעיית התכנות הליניארי (3.2) ומציאת החסמים עבור משפחה זו נעשה ע"י פתרון בעיית התכנות הליניארי הבאה:

$$\begin{aligned}
&\min_P \Big/ \max_P \ \sum_{i=0}^{N-1} P_{ii} \cdot \left(\pi Q_N \pi^T D\right)_{ii} \\
&\text{s.t.} \quad \sum_{i=0}^{N-1} P_{ii} \cdot f_k\left(\underline{\phi}_i\right) \le M_k \quad k = 0,1,\ldots,K-1 \\
&\sum_{i=0}^{N-1} P_{ii} = 1 \\
&P_{ii} \ge 0
\end{aligned} \tag{3.5}$$

ניתן למשל להוסיף אילוץ על ממוצע אפס: $\sum_{i=0}^{N-1} P_{ii} \cdot \underline{\phi}_i = 0$. אילוץ זה לא הביא להידוק החסמים, ברב המקרים, מכיוון שהחסם התחתון ממומש ע"י וקטור ייצוג הקרוב לאפס והחסם העליון ממומש ע"י מקור המרוכז בשתי נקודות קצה מנוגדות כך שהממוצע הוא אפס.

### 3.3.1 שיפור החסמים ע"י שימוש בהגבלת הספק

דוגמה חשובה לשימוש במידע על מומנטים הוא הגבלת ההספק של המקורות. נסתכל, לדוגמה, על קוונטייזר סקלרי אחיד בתחום $[-1,1]$, כאשר רמות הייצוג מסומנות (בסדר עולה) ב-$\phi_k$ , $k = 0,1,\ldots,N-1$. המרחק בין רמות הייצוג יסומן ב-$\Delta$ (צעד הקוונטיזציה). הקוונטייזר מוזן ממקור בעל ממוצע אפס ושונות המנורמלת ליחידה. ניתן לחלק את שגיאת הקוונטיזציה הכוללת לשני גורמים [Ja84 סעיף 4.2]:

א. <u>רעש גרגירי</u> (Granular Noise) - כאשר דגימה $x$ נמצאת בתוך "תחום" הקוונטייזר, $\phi_0 - \frac{\Delta}{2} \le x \le \phi_{N-1} + \frac{\Delta}{2}$. זהו העיוות הנגרם כתוצאה מן המרחק בין הדגימה לבין רמת הייצוג הקרובה. מרחק זה מוגבל לחצי המרחק בין רמות ייצוג שכנות.

ב. <u>רעש "העמסה"</u> (Overload Noise) - כאשר דגימה $x$ איננה נמצאת בתוך "תחום" הקוונטייזר, העיוות נגרם כתוצאה מן המרחק בין הדגימה לבין **רמת הייצוג הקיצונית** הקרובה יותר ($\phi_k$ , $k = 0$ or $N-1$). מרחק זה איננו מוגבל.

בחירת צעד קוונטיזציה $\Delta$ קטן, תגרום להקטנת המרחק בין רמות-ייצוג שכנות ולכן תקטין את הרעש הגרגירי. מאידך יותר דגימות תופענה מחוץ לתחום הקוונטייזר ורעש ה-"העמסה" יגבר.

מאידך, בחירת צעד קוונטיזציה $\Delta$ גדול תגדיל את הרעש הגרגירי ותקטין את רעש הקצה.

קיימת נקודה $\Delta_{opt}$ בה עיוות הקוונטיזציה הכולל המורכב משני הגורמים מגיע למינימום. את היחס בין תחום הקוונטייזר לסטיית התקן של האות $\sigma_x$, מקובל לכנות גורם גלישה (Overload factor) - $\gamma = \dfrac{N\Delta}{\sigma_x}$.

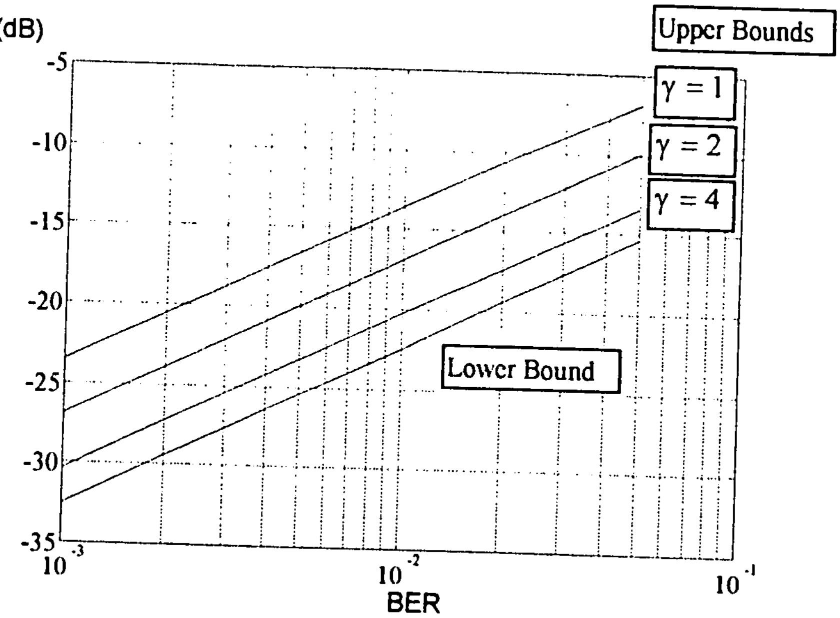
עבור קוונטייזרים סקלריים לשימושים של קידוד דיבור מקובל להשתמש בגורם גלישה - $\gamma$ המבטיח כי לא תיגרם גלישה בתחום של 3-4 סטיות תקן.

$$\sum_{i=0}^{N-1} P_{ii} \cdot \phi_i^2 \le \frac{2}{\gamma^2} \qquad \gamma \approx 3 \div 4 \tag{3.6}$$

הוספת אילוץ זה לבעיית התכנות הליניארית תביא לבעיה בת 2 משוואות ב-$N$ נעלמים. במקרה זה ישנם $\binom{N+1}{2}$ פתרונות בסיסיים, לא כולם קבילים (Feasible), לבעיית התכנות הליניארי. יש לחשב את ערך פונקצית המטרה ב-(3.5) עבור פתרונות אלו, ולאתר את הגבוה ואת הנמוך ביותר.

כפי שראינו בציור 3.1 אין משמעות לחפש חסמים הדוקים יותר עבור הקוונטייזר הסקלרי האחיד עם השיבוץ הבינארי הטבעי. מסיבה זו נשתמש כאן בקוד Gray, כדי להדגים את שיפור החסמים ע"י הוספת אילוץ של הגבלת הספק.

בציור 3.4 מתוארים החסמים לקוונטייזר סקלרי אחיד בן 16 רמות בתחום [-1,1], כאשר הפעם, כאמור, נעשה שימוש בקוד Gray. החסמים, בניגוד לשיבוץ הבינארי הטבעי, אינם מתלכדים. בציור מתוארים החסמים עבור גורמי גלישה $\gamma = 1,2,4$. גורם גלישה $\gamma = 1$ מביא לחסם הבסיסי.




ציור 3.4 - חסמים על עיוות הערוץ מעל משפחות מקורות בעלות גורמי גלישה נתונים (קוונטייזר סקלרי אחיד 4 סיביות בתחום $[-1,1]$, ערוץ בינארי סימטרי, קוד Gray)

Fig 3.4 - Bounds on Channel-Distortion over signal sources with specific overload factors. (4 bit Uniform Scalar Quantizer covering $[-1,1]$, Binary Symmetric Channel, Gray Code)

החסם התחתון משותף לכל גורמי הגלישה והוא ממומש ע"י מקור המרוכז באפס. מן החסם התחתון נראה כי עבור מקורות "טובים" הקוד נותן ביצועים הטובים בערך ב-4dB מביצועי הקוד הבינארי (מתוך השוואה עם ציור 3.1). החסמים העליונים ממומשים במקורות המרוכזים בקצות התחום המותר ע"י הגבלת ההספק. הגדלת גורם הגלישה (ובכך הקטנת הספק האות) מביאות להורדת החסם העליון. עבור גורם גלישה $\gamma = 4$, גם החסם העליון נמוך מן העיוות הנגרם עבור קוד בינארי טבעי. מכאן ששיבוץ בקוד Gray עדיף על שיבוץ בקוד הבינארי הטבעי עבור מקורות המקיימים את מגבלת ההספק המתאימה ל-$\gamma \geq 3$. בפרק 5 נראה כי עבור מקור אחיד ביצועי השיבוץ הבינארי הטבעי הם העדיפים.

## 3.4 חסימת עיוות הערוץ עבור קצבי שגיאות קטנים

נסתכל על הערוץ הבינארי הסימטרי או על ערוצים הנובעים ממנו ע"י קוד בלוק לתיקון שגיאות כדוגמת קוד Hamming שתואר בציור 3.2. איברי מטריצת הערוץ $Q$ תלויים בהסתברות השגיאה בערוץ הבינארי הסימטרי - $q$. נסמן עובדה זו באופן מפורש ע"י $Q(q)$.

עבור קצבי שגיאות קטנים אברי האלכסון במטריצת הערוץ $Q$ הם הדומיננטיים. מאידך, מכיוון שהמרחק בין ווקטור ייצוג לעצמו הוא תמיד אפס, איברים אלו מוכפלים באפסים וניתן לאפס את האלכסון של מטריצת הערוץ $Q$.

נגדיר:

$$\tilde{Q}_{ij}(q)=\begin{cases}Q_{ij}(q) & i\neq j\\ 0 & i=j\end{cases} \qquad (3.7)$$

כעת נחפש את גורם השגיאה הדומיננטי בשאר איברי מטריצת הערוץ. המטריצה $\tilde{Q}(q)$ מורכבת מסכומים של איברים מן הצורה $l=L^0,L^0+1,\ldots,L$ $q^l(1-q)^{L-l}$, כאשר סימנו את החזקה הנמוכה ביותר ב-$L^0$. הסיבה לכך שאיפסנו את האלכסון של המטריצה המקורית $Q(q)$היא שהיו באלכסון איברים בחזקות נמוכות יותר מ-$L^0$. נבחן את הגבול:

$$\lim_{q\to 0}\frac{1}{q^{L^0}}\cdot\tilde{Q}(q)=\tilde{Q}_A \qquad (3.8)$$

איברים עם חזקות של $q$ הגבוהות מ-$L^0$ יתאפסו בגבול.
הגבול של עיוות הערוץ עבור קצבי שגיאה נמוכים בערוץ הוא לכן:

$$\lim_{q\to 0}\frac{1}{q^{L^0}}D_c=trace\left\{P\pi\tilde{Q}_A\pi^T D\right\} \qquad (3.9)$$

ניתן בצורה זו למצוא את ההתנהגות האסימפטוטית של עיוות הערוץ.
עבור הסתברויות שגיאה קטנות (מתחת לאחוז אחד) ההשפעה של גורם מסוג $q^l(1-q)^{L-l}$ גדולה בלפחות 20dB מהשפעתו של גורם מסוג $q^{l+1}(1-q)^{L-l-1}$. בהמשך ניווכח לדעת כי ההתנהגות האסימפטוטית קרובה מאד להתנהגות המדויקת אפילו עבור הסתברויות שגיאה גבוהות יחסית של 5 אחוז.
ניתן להיעזר בטכניקות החסימה (מעל כל המקורות האפשריים) שתוארו בסעיף 3.2 גם לקבלת ההתנהגות האסימפטוטית של החסמים על עיוות הערוץ עבור קצבי שגיאות קטנים $(q\to 0)$. ניתן גם לשפר את החסמים בעזרת המידע על מומנטים כפי שנעשה בסעיף 3.3.
לדוגמה, עבור הערוץ הבינארי הסימטרי קצב השגיאה הדומיננטי הוא $L^0=1$. החסמים יהיו על הגודל: $\lim_{q\to 0}\frac{1}{q}D_c$.
בציור 3.1 ובציור 3.4 תיארנו את הביצועים של קוונטיזר אחיד בן 4 סיביות בתחום [-1,1] עבור קוד בינארי סימטרי ועבור קוד Gray בהתאמה.

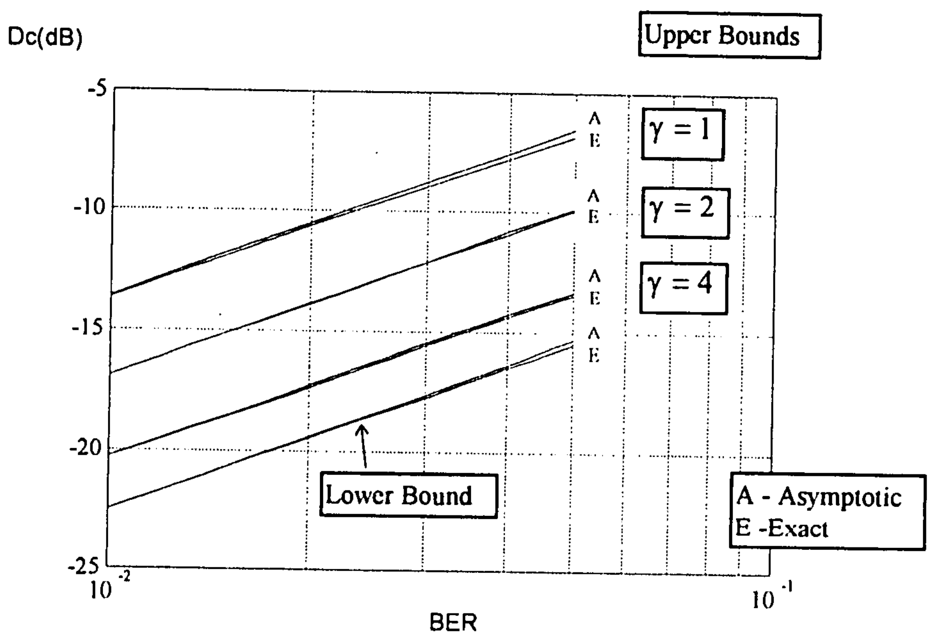
במקרים אלו נתקבלו הקבועים הבאים:

| Code | | Asymptotic Lower Bound | Asymptotic Upper Bound |
|---|---|---|---|
| Natural Binary | | 1.33 | 1.33 |
| Gray | $\gamma = 1$ | 0.56 | 4.43 |
| | $\gamma = 2$ | 0.56 | 2.06 |
| | $\gamma = 4$ | 0.56 | 0.94 |

טבלה 3.1 - חסמים על הערכים האסימפטוטיים של עיוות הערוץ המנורמל $(D_c/q)$ מעל כל המקורות האפשריים, ומעל מקורות עם הגבלת הספק (קוונטייזר אחיד 4 סיבית בתחום $[-1,1]$, ערוץ בינארי סימטרי)

Table 3.1 - Bound on the asymptotic Normalized Channel-Distortion $(D_c/q)$ over all possible signal sources and power limited source (4 bit Uniform Quantizer range $[-1,1]$, Binary Symmetric Channel)

בציור 3.5 מתוארים הערכים האסימפטוטיים בהשוואה לחסמים המקוריים עבור קוד Gray מציור 3.4.




ציור 3.5 - חסמים על עיוות הערוץ מעל משפחות מקורות בעלות גורמי גלישה נתונים (קוונטייזר סקלרי אחיד 4 סיבית בתחום $[-1,1]$, ערוץ בינארי סימטרי, קוד Gray) בהשוואה לחסמים אסימפטוטיים

Fig 3.5 - Bounds on Channel-Distortion over signal sources with specific overload factors. (4 bit Uniform Scalar Quantizer range $[-1,1]$, Binary Symmetric Channel, Gray Code) Compared with Asymptotic bounds

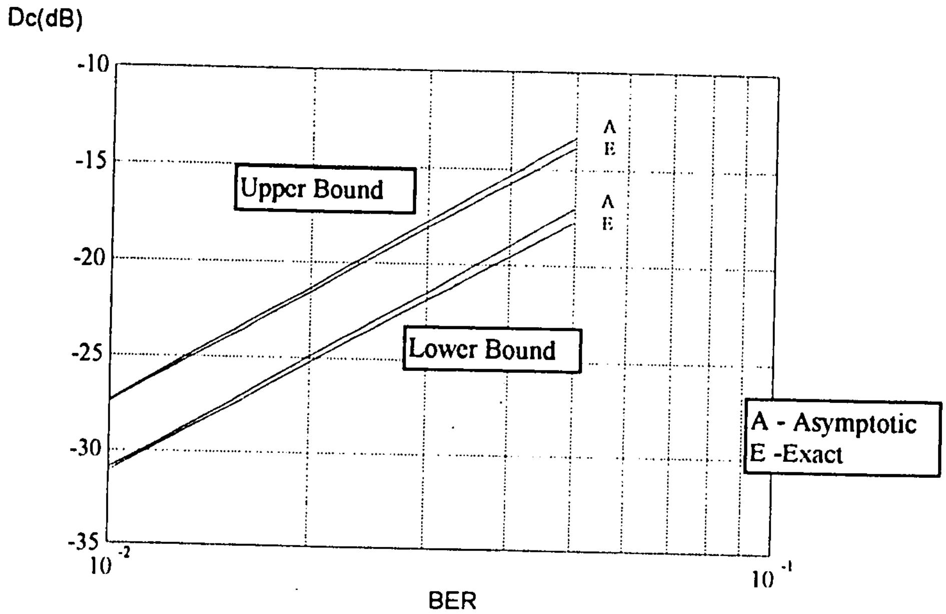
ניתן לראות כי ההתנהגות האסימפטוטית קרובה מאד להתנהגות האמיתית עד להסתברות שגיאה של כ-5 אחוז.

דוגמה נוספת: עבור הערוץ הבינארי הסימטרי עם קוד Hamming (7,4) לתיקון שגיאות, קצב השגיאה הדומיננטי הוא $L^0 = 2$.

עבור קוונטיזר אחיד בן 4 סיביות בתחום [-1,1] וקוד תיקון זה (הופיע בציור 3.2), ההתנהגות האסימפטוטית היא:

$$\sim 7.98 \le \lim_{q\to 0} \frac{1}{q^2} D_c \le \sim 18.45 \tag{3.10}$$

בציור 3.6 מתוארים הערכים האסימפטוטיים בהשוואה לחסמים המקוריים.




ציור 3.6 - חסמים על עיוות הערוץ מעל כל מקורות האות האפשריים עבור קוונטייזר סקלרי אחיד 4 סיביות, ערוץ בינארי סימטרי וקוד Hamming (7,4) בהשוואה לחסמים אסימפטוטיים

Fig. 3.6 - Bounds on Channel-Distortion over all possible signal sources for 4-bit Uniform Scalar Quantizer, Binary Symmetric Channel and (7,4) Hamming Code. The exact bound are compared with asymptotic bounds

גם במקרה זה ניתן לראות כי ההתנהגות האסיפטוטית קרובה מאד להתנהגות המדויקת עד להסתברויות שגיאה של כ-5 אחוז.

## 3.5 הקוונטייזר הסקלרי האחיד עם הקוד הבינארי הטבעי תחת ערוץ בינארי סימטרי

בסעיף זה נסקור מספר תכונות מעניינות המתקבלות במקרה המיוחד של קווניטייזר סקלרי אחיד וקוד בינארי טבעי (NBC) תחת הערוץ הבינארי הסימטרי (BSC).

מטבלה 3.1 ומציור 3.1 ניתן לראות כי החסם העליון והתחתון על עיוות הערוץ מתלכדים עבור הסתברויות שגיאה בערוץ הקטנות מחמישה אחוז. המסקנה היא שביצועי הקוונטייזר הסקלרי

האחיד במקרה זה **אינם תלויים בהסתברויות של מקור האות**. כאמור, הסיבה לכך היא שעבור הקוד הבינארי הטבעי לכל סיבית ניתן להתאים עיוות ריבועי אוקלידי. לספרה המשמעותית ביותר יש ערך יחסי השווה $1^2$, לסיבית השנייה ערך $\frac{1}{2^2}$ וכו'. עבור הסתברויות שגיאה נמוכות, מאורע השגיאה השכיח של שגיאה בסיבית אחת ייתן בהסתברות שווה עיוות לפי ערכים אלו, ללא תלות במקור.

בטבלה 3.2 מופיעים החסמים האסימפטוטיים עבור קוונטייזרים סקלריים אחידים עבור מספר סיביות שונה.

| Number of bits | Asymptotic Upper and Lower Bounds $\lim_{q\to 0} D_c/q$ |
|---|---|
| 2 | 1.250 |
| 3 | 1.313 |
| 4 | 1.328 |
| 5 | 1.332 |
| 6 | 1.333 |
| 7 | 1.333 |
| 8 | 1.333 |

טבלה 3.2 - חסמים אסימפטוטיים על עיוות הערוץ עבור קוונטייזרים סקלריים אחידים בתחום $[-1,1]$ עם מספר סיביות שונה, שיבוץ בינארי טבעי תחת ערוץ בינארי סימטרי

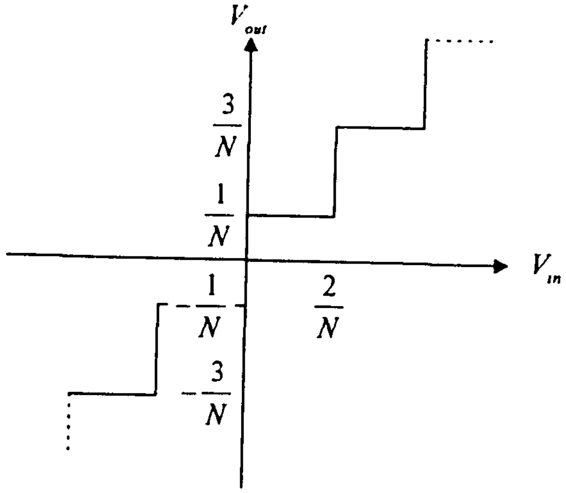
Table 3.2 - Asymptotic bounds on the channel distortion for scalar quantizers in the range $[-1,1]$ with various number of bits,

Natural Binary Code and Binary Symmetric Channel

מן הטבלה ניתן לראות כי עיוות הערוץ במקרה זה כמעט ואיננו תלוי במספר הסיביות $L$ של הקוונטייזר הסקלרי. הסיבה לכך היא שסכום הערכים : $\sum_{l=0}^{L-1}\frac{1}{\left(2^l\right)^2}$ מתכנס במהירות לערך $\frac{4}{3}$.

נקדים את המאוחר ונציין כי בפרק 5 מופיעה הוכחה כי **השיבוץ הבינארי הטבעי הוא השיבוץ האופטימלי** עבור <u>קוונטייזר סקלרי אחיד, מקור אחיד וערוץ בינארי סימטרי</u>. כאן נראה כי <u>עבור כל מקור שהוא, השיבוץ הבינארי הטבעי טוב מן הממוצע על כל השיבוצים האפשריים</u>.

ממוצע העיוות (Ensemble Average) - $\overline{D}_c$ הוא הממוצע של עיוות הערוץ על פני כל השיבוצים האפשריים של מילות קוד לוקטורי-ייצוג (סה"כ $N!$). ב-[Fa90] מופיע חישוב אסימפטוטי של ממוצע העיוות עבור הערוץ הבינארי הסימטרי. במקרה זה ניתן להגיע לביטוי פשוט לממוצע העיוות (מוצג בהמשך). רמות הייצוג מתוארות בציור 3.7.




ציור 3.7 - רמות הייצוג של קוונטייזר אחיד סקלרי מסוג Midrise

Fig. 3.7 - Representing levels of a Midrise scalar uniform quantizer

נתאר את רמות הייצוג בוקטור עמודה:

$$\underline{\rho}=\begin{bmatrix}\rho_0\\ \rho_1\\ \vdots\\ \rho_i\\ \vdots\\ \rho_{N-1}\end{bmatrix}=\frac{2}{N}\cdot\begin{bmatrix}-N/2+1/2\\ 1-N/2+1/2\\ \vdots\\ i-N/2+1/2\\ \vdots\\ N/2-1+1/2\end{bmatrix} \tag{3.11}$$

חישוב ממוצע העיוות נותן:

$$\lim_{q\to 0}\frac{\overline{D}_c}{q}=\lim_{q\to 0}\frac{1}{q}\left(\frac{1}{N!}\sum_{\text{Every assignment } \pi} D_c(\pi)\right)=$$

$\Leftarrow$ [Fa90]

$$=\frac{L}{N-1}\sum_{i=0}^{N-1}P_i\sum_{j=0}^{N-1}\left(\rho_i-\rho_j\right)^2=$$

$$=\frac{(4N-2)L}{3N}-\frac{4L}{N(N-1)}\sum_{i=0}^{N-1}P_{ii}\left[(N-1)i-i^2\right] \tag{3.12}$$

האיבר השני בסכום כולל ממוצע משוקלל של איברי "פרבולה" בדידה. ניתן לחסום את ממוצע העיוות ע"י הצבת הערך הגבוה והערך הנמוך ב"פרבולה" זו. מכאן מתקבל:

$$\left(\lim_{q\to 0}\frac{\overline{D}_c}{q}\right)_{min} = \frac{L}{3N}\left[N+1+\frac{3}{N-1}\right] \approx \frac{1}{3}L$$
$$\left(\lim_{q\to 0}\frac{\overline{D}_c}{q}\right)_{max} = \frac{(4N-2)L}{3N} \approx \frac{4}{3}L \qquad (3.13)$$

כאשר $L$ הוא מספר הסיביות ו-$N = 2^L$ מספר הרמות בקוונטייזר הסקלרי האחיד.

בטבלה 3.3 מופיעים ערכי המינימום והמקסימום של ממוצע העיוות האסימפטוטי (מנורמל להסתברות השגיאה בערוץ) בהשוואה לעיוות הנגרם עבור הקוד הבינארי הטבעי, עבור קוונטייזרים סקלריים עם מספר שונה של סיביות.

| No. of bits ($L$) | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| Natural Binary Code | 1.25 | 1.31 | 1.33 | 1.33 | 1.33 | 1.33 | 1.33 |
| Ensemble Av. Lower Bound | 1.00 | 1.18 | 1.43 | 1.72 | 2.03 | 2.35 | 2.68 |
| Ensemble Av. Upper Bound | 2.33 | 3.75 | 5.17 | 6.56 | 7.94 | 9.30 | 10.65 |

טבלה 3.3 - חסמים אסימפטוטיים מעל כל המקורות האפשריים על ממוצע עיוות הערוץ על פני כל השיבוצים האפשריים ($\overline{D}_c/q$), בהשוואה לביצועי הקוד הבינארי הטבעי עבור קוונטייזרים סקלריים אחידים עם מספר סיביות שונה

Table 3.3 - Asymptotic bounds over all possible source probabilities on the Channel Distortion Ensemble Average over all possible assignments ($\overline{D}_c/q$), compared with the performance of the Natural Binary Code assignment.

מן הטבלה ניתן להבחין כי ממוצע העיוות עולה עם מספר הסיביות $L$, בעוד שהקוד הבינארי הטבעי שומר על עיוות כמעט קבוע. עבור קוונטייזרים בני ארבע סיביות ויותר, הקוד הבינארי הטבעי טוב אפילו מן החסם התחתון על ממוצע העיוות ולכן טוב מממוצע העיוות עבור כל הסתברויות המקור האפשריות. תכונה זו מתאימה לאובזרבציה של [Fa90] אשר מצא אמפירית כי עבור קוונטייזר וקטורי במבנה עץ, השיבוץ הטבעי נותן עיוות הנמוך מן הממוצע על עיוות הערוץ על פני כל השיבוצים האפשריים.

## 3.6 סיכום

בפרק זה קיבלנו חסמים על עיוות הערוץ מעל כל המקורות האפשריים המזינים את הקוונטייזר. הצגנו דוגמאות מספריות שונות עבור ערוץ בינארי סימטרי בלי ועם קוד לתיקון שגיאה בסיבית אחת. ראינו כי במקרים טיפוסיים החסם התחתון ממומש ע"י מקור המרוכז בנקודה אחת בסביבת האפס, כאשר החסם העליון ממומש ע"י מקור המרוכז בתא קצה. מסיבה זו הצלחנו לקבל חסמים הדוקים יותר כאשר צמצמנו את משפחת המקורות ע"י תוספת אילוצי-מומנטים ובעיקר הגבלת הספק. ראינו דרך למציאת חסמים על ההתנהגות האסימפטוטית של עיוות הערוץ עבור הסתברויות שגיאה קטנות. נוכחנו לראות כי בדוגמאות שהוצגו ההתנהגות האסימפטוטית קרובה להתנהגות המדויקת גם בהסתברויות שגיאה בערוץ של עד כחמישה אחוז.
עבור המקרה המיוחד של קוונטייזר סקלרי אחיד ושיבוץ בינארי סימטרי ראינו כי עיוות הערוץ כמעט ואיננו תלוי במקור וכמעט ואיננו תלוי במספר הסיביות של הקוונטייזר. ראינו גם כי השיבוץ הבינארי הטבעי נותן עיוות ערוץ נמוך יותר מממוצע עיוות הערוץ על פני כל השיבוצים האפשריים (Ensemble Average) עבור כל מקור שהוא, עבור 4 סיביות ויותר.

# פרק 4 - כיווץ ליניארי של קוונטייזרים וקטוריים להקטנת העיוות בתנאי שגיאות ערוץ

## 4.1 מבוא

בסקר הספרות תיארנו את התנאים ההכרחיים לקוונטייזר וקטורי אופטימלי הפועל תחת שגיאות ערוץ בשיבוץ אינדקסים נתון. ראינו כי תנאי השכן הקרוב ותנאי הצנטרואיד הוחלפו בתנאים דומים המשקללים את השפעת שגיאות הערוץ. קוונטייזר מסוג זה נקרא בספרות [Kn92] בשם COVQ - Channel Optimized Vector Quantizer.

תכן קוונטייזר מסוג COVQ הוא קשה ומסובך הרבה יותר מתכן שאיננו מתחשב בשגיאות ערוץ. תוצאת התכן היא משפחה של קוונטייזרים וקטורים (ספרי-קוד של וקטורי ייצוג) המתאימים כל אחד להסתברויות שגיאה (BER - Bit Error Rate) אחרות בערוץ. מסיבה זו המימוש של COVQ הוא יקר יחסית בדרישות זיכרון הן במשדר והן במקלט.

בפרק זה מוצע מבנה פשוט של קוונטייזר וקטורי להקטנה תת-אופטימלית של השפעת שגיאות הערוץ. השינוי המוצע הוא כיווץ ליניארי של מבנה הקוונטייזר (שתוכנן ללא התחשבות בשגיאות ערוץ) אותו נכנה - SVQ - Scaled Vector Quantizer.

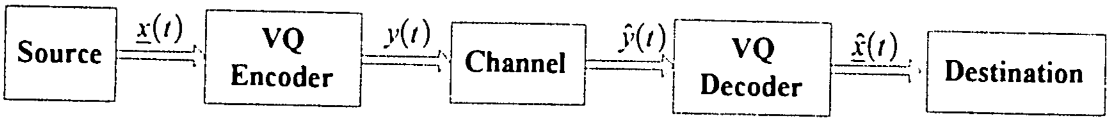
תכן הקוונטייזר נעשה בשני שלבים: בשלב ראשון תכן קוונטייזר ללא התחשבות בשגיאות הערוץ (NVQ - Noiseless Vector Quantizer) ובשלב שני מציאת גורמי הכיווץ המתאימים לקצבי שגיאות שונים בערוץ. גם המימוש הוא פשוט בהרבה יחסית ל-COVQ. מספיק לשמור את ספר הקוד של ה-NVQ וטבלה של סקלרים המתאימים לערכי BER שונים.

בדוגמאות מספריות אנו מראים ביצועי ה-SVQ נופלים במידה "סבירה" מביצועי ה-COVQ בהתחשב בפישוט התכן והמימוש.

## 4.2 תיאור הכיווץ הליניארי בקוונטייזר וקטורי

מערכת תקשורת המבוססת על קוונטייזר וקטורי המתוכנן ללא התחשבות בשגיאות ערוץ מתוארת בציור 4.1. בהמשך נקרא לסוג זה של קוונטייזר בשם NVQ - Noiseless Vector Quantizer. תכן הקוונטייזר נעשה בשיטות ה-"קלסיות" המקובלות.

כזכור, התנאים ההכרחיים לאופטימום הם תנאי השכן הקרוב (2.8) ותנאי הצנטרואיד (2.9). אלגוריתמים לתכן NVQ כדוגמת LBG או Generalized-Lloyd (ראה [Ge92, סעיף 11.3]) מבוססים על יישום לסירוגין של התנאים ההכרחיים על סדרות אימון (או על הסתברות המקור, כאשר היא ידועה). שיטת תכן זו מבטיחה התכנסות למינימום לוקלי. כאשר מופעל NVQ תחת שגיאות ערוץ נגרמת הרעה בביצועי המערכת כפי שתוארה בסקר הספרות (סעיף 2.2).

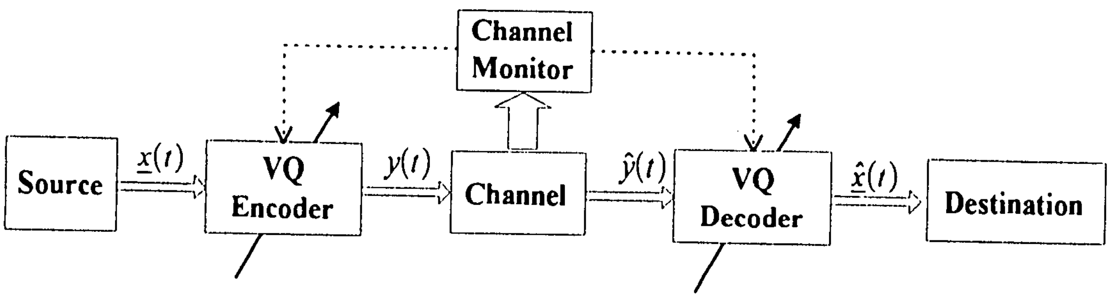



ציור 4.1 - מערכת קוונטיזציה וקטורית

Fig. 4.1 - Vector Quantization System

כאמור, בספרות הוצעו תנאים הכרחיים אלטרנטיביים לאופטימום של קוונטייזר וקטורי תחת שגיאות ערוץ - [Ku69], [Ze88], [Fa90] ו-[Fa91]. התנאים ההכרחיים הינם תנאי השכן הקרוב המשוקלל (2.22) ותנאי הצנטרואיד המשוקלל (2.23). בניגוד למקרה בו לא קיימות שגיאות ערוץ, יישום לסירוגין של שני התנאים איננו מבטיח התכנסות. כמו-כן לא מובטחת רגולריות במבנה הקוונטייזר.

בפועל מתכננים בשלב ראשון קוונטייזר ללא התחשבות בשגיאות הערוץ - NVQ. בהמשך מריצים את אלגוריתם התכן עבור קצב שגיאות נמוך, כאשר ה-NVQ מהווה את תנאי ההתחלה לתכן. בהמשך, בכל שלב בתכן מעלים בהדרגה את גורם קצב שגיאות הערוץ בתנאים (2.22) ו-(2.23), כאשר תנאי ההתחלה הם תוצאות שלב התכן הקודם. בצורה זו מקבלים סדרה של קוונטייזרים שונים המותאמים כל אחד לקצב שגיאות שונה בערוץ. לקוונטייזר שתוכנן בדרך זו נקרא COVQ - Channel Optimized Vector Quantizer ומערכת תקשורת מתאימה מתוארת בציור 4.2.




ציור 4.2 - מערכת קוונטיזציה וקטורית מותאמת לערוץ

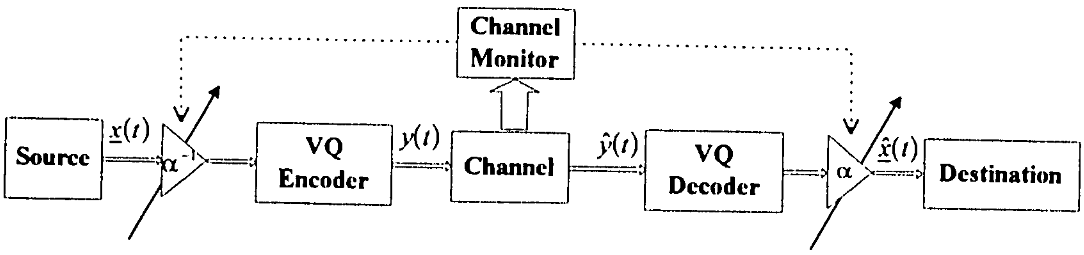
Fig. 4.2 - Channel Optimized Vector Quantization System

מכיון שתכן ה-COVQ תלוי בקצב השגיאות בערוץ, הרי שגם המשדר וגם המקלט צריכים לדעת את תנאי הערוץ. לשם כך יש להוסיף למערכת מודול ניטור ערוץ (Channel Monitor). מודול הניטור מודיע למשדר ולמקלט על תנאי הערוץ ושניהם בוחרים את הקוונטיזר המתאים לקצב השגיאות הנתון.

החיסרון הבולט במימוש של גישה זו הוא הצורך בקוונטייזרים שונים עבור תנאים שונים של שגיאות הערוץ. דרישה זו באה לידי ביטוי בדרישות זיכרון במקלט ובמשדר לשמירת סדרת הקוונטייזרים.

בתכן קוונטייזרים המותאמים לערוץ (COVQ) ניתן להבחין בהצטופפות וקטורי הייצוג סביב "מרכז הכובד" של הסתברות המקור. להצטופפות וקטורי הייצוג ישנם שתי השפעות מנוגדות. מצד אחד כל זוג של וקטורי ייצוג מתקרבים זה לזה ולכן שגיאת ערוץ תביא לעיוות קטן יותר. מצד שני הקוונטייזר מתרחק מן התכן ללא שגיאות ערוץ ולכן נפגעת התאמתו למקור. בפרט, רעש "ההעמסה" (Overload) יגדל עם כיווץ הקוונטייזר. ככל שיגבר קצב השגיאות בערוץ, תגדל השפעתו של גורם שגיאות הערוץ והקוונטייזר יהיה צפוף יותר.

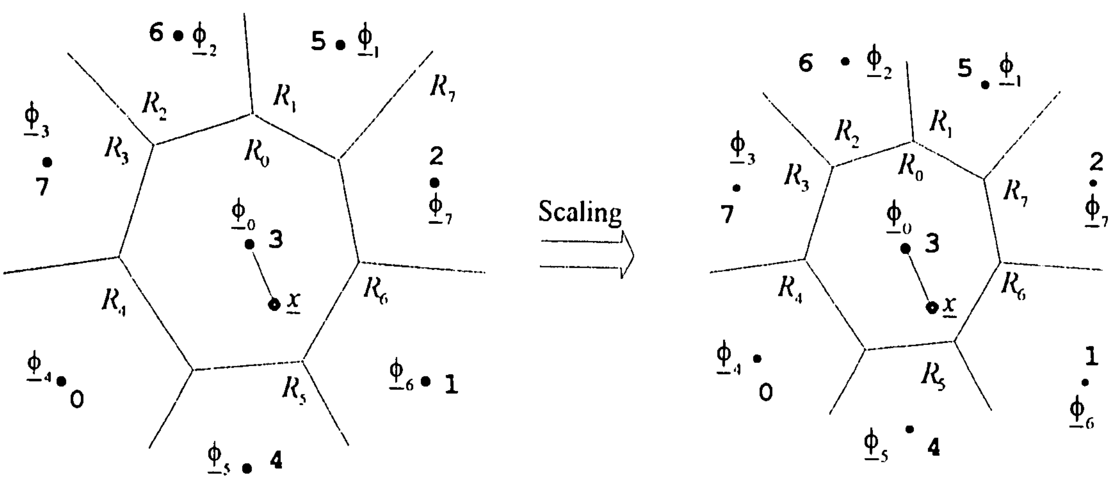
הבחנה זו נתנה את המוטיבציה לדרך ביניים להקטנת השפעת שגיאות הערוץ והיא כיווץ ליניארי של מבנה הקוונטייזר - Scaled Vector Quantizer - SVQ המתוארת בציור 4.3.




ציור 4.3 - מערכת קוונטיזציה וקטורית עם כיווץ ליניארי

Fig. 4.3 - Scaled Vector Quantization System

המקודד והמפענח במערכת ה-SVQ זהים לקוונטייזר שתוכנן ללא התחשבות בשגיאות הערוץ - NVQ. השינוי להקטנת השפעתן של שגיאות הערוץ הוא כפל בסקלר $1/\alpha$ במבוא המקודד וכפל בסקלר $\alpha$ במוצא המפענח. הסקלר יכול לקבל ערכים בתחום $0 < \alpha \le 1$. בתאור המערכת הנחנו ממוצע אפס של אות המקור כך שכפל בסקלר שקול להצטופפות סביב מרכז ההסתברות. בציור 4.4 מתוארים השינויים כתוצאה מן הכפל בסקלר:




ציור 4.4 - תאור סכימטי של קוונטייזר וקטורי עם כיווץ ליניארי

Fig. 4.4 - Schematic drawing of a Scaled Vector Quantizer

הכפל במבוא המקודד בסקלר $\alpha^{-1}$ גורם לכך שלמקודד מוצג הוקטור ה"מוגדל" $\alpha^{-1}\underline{x}(t)$.
מנקודת מבטו של המקור היה עליו להפיק את הוקטור $\alpha\underline{x}(t)$ על מנת לשמור את מבנה תאי החלוקה בקוונטייזר. לפיכך הכפל במבוא המקודד שקול לכיווץ בפקטור $\alpha$ של תאי החלוקה.
הכפל במוצא המפענח בסקלר $\alpha$ שקול לכיווץ הוקטורים המייצגים.
עבור מצבים שונים של שגיאות בערוץ ניתן לחפש Off-Line את הערך האופטימלי של הסקלר $\alpha$.
במימוש אין צורך להחזיק ספרייה של קוונטייזרים וקטוריים כדוגמת ה-COVQ. במקום זאת, מספיק לשמור במשדר ובמקלט מימוש של קוונטייזר יחיד ובנוסף טבלה (Look-Up-Table) המחזיקה ערכים אופטימליים של הסקלר $\alpha$. בהמשך הפרק נדון בביצועים של SVQ ונציג מספר דוגמאות מספריות.

## 4.3 קוונטייזר וקטורי עם כיווץ ליניארי (Scaled Vector Quantizer)

נסתכל על קוונטייזר וקטורי עם כיווץ ליניארי הפועל בתנאי ערוץ בינארי סימטרי. בפרק זה נשתמש במידת מרחק ריבועית. כאמור, הסקלר $\alpha$ גורם לכיווץ תאי החלוקה ולהצטופפות של וקטורי הייצוג. העיוות הכולל (2.17) עבור מקרה זה הוא:

$$D_{\text{Total}}(\alpha)=\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\pi Q\pi^{T}\right)_{ij}\int_{R_i(\alpha)}\left\|\underline{x}-\alpha\underline{\phi}_j\right\|^2\cdot f_{\underline{x}}(\underline{x})\cdot d\underline{x} \tag{4.1}$$

כאשר תא החלוקה החדש מוגדר ע"י $\underline{x}\in R_i \Leftrightarrow \alpha\underline{x}\in R_i(\alpha)$ וכן $R_i(1)=R_i$. עבור הערוץ הבינארי הסימטרי $Q_{ij}=q^{H(i,j)}(1-q)^{L-H(i,j)}$ כאשר $H(i,j)$ הוא מרחק Hamming בין הייצוגים הבינאריים של המספרים $i$ ו- $j$ ו- $L=log_2 N$ הוא מספר הסיביות.
עבור קוונטייזר ומקור נתונים ועבור הסתברות שגיאה נתונה בערוץ $q$, מידת המרחק הכוללת $D_{\text{Total}}(\alpha)$ היא פונקציה במשתנה אחד $\alpha$. עבור הסתברויות שגיאה שונות ניתן לחשב Off-Line את גורם הכיווץ האופטימלי $\alpha_{opt}(q)$. במימוש המערכת ניתן לאכסן את $\alpha_{opt}(q)$ בטבלה אשר תימצא גם במשדר וגם במקלט. בדוגמאות המספריות המוצגות בהמשך הפרק נראה כי במקרים שנבחנו, הפונקציה $\alpha_{opt}(q)$ היא פשוטה לתיאור. יתרה מזו, לפונקצית מידת המרחק הכוללת $D_{\text{Total}}(\alpha)$ קיים בדוגמאות מינימום שטוח כך שידיעת הסתברות השגיאה המדויקת איננה הכרחית. לחילופין, אין צורך בידיעה מדויקת של $\alpha_{opt}(q)$.
עבור $\alpha=1$ מתקבל העיוות עבור הקוונטייזר שתוכנן ללא התחשבות בשגיאות ערוץ - NVQ. לכן ללא שגיאות בערוץ $q=0$, המינימום של $D_{\text{Total}}(\alpha)$ יימצא ב - $\alpha=1$.

כללית, לשינוי ב- $\alpha$ ישנן שתי השלכות:

א. תזווה מן התכן המקורי המתחשב במקור בלבד (ולא בשגיאות הערוץ), ה-NVQ. נצפה כי הגדלת $\alpha$ או הקטנתו תגדיל את עיוות הקוונטיזציה.

ב. הקטנת $\alpha$ תגרום לכל זוג וקטורי ייצוג להתקרב זה לזה. השפעתו של העיוות כתוצאה משגיאה בערוץ תקטן. הגדלת $\alpha$ תגדיל בהתאם את עיוות הערוץ.

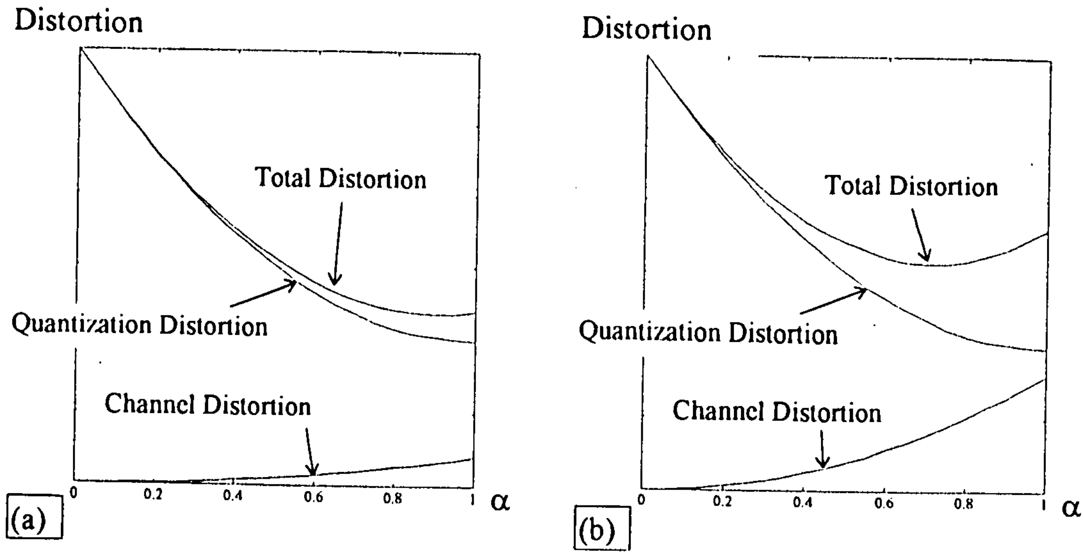
ההתנהגות הטיפוסית של הפונקציה $D_{\text{Total}}(\alpha)$ תהיה:

א. הגדלת $\alpha$ לערך הגדול מ-1 תגרום להגדלת העיוות הכולל.

ב. לפונקציה יש מינימום בתחום $0 < \alpha \le 1$.

ג. מיקומו של המינימום יהיה קרוב ל-1 עבור הסתברויות שגיאה קטנות בערוץ והוא יתקרב לאפס עם הגדלת קצב השגיאות בערוץ.

תיאור סכימטי של העיוותים במערכת בתלות בסקלר $\alpha$ עבור עיוות ערוץ נמוך ועבור עיוות ערוץ גבוה מופיע בציור 4.5:




ציור 4.5 - תיאור סכימטי של עיוותים טיפוסיים כפונקציה של גורם הכיווץ $\alpha$

(a) - עיוות ערוץ נמוך, (b) - עיוות ערוץ גבוה

Fig 4.5 - Schematic drawing of typical distortion as a function of the scaling factor $\alpha$

(a) - Low Channel Disotortion, (b) - Severe Channel Distortion

מן הסכימה ניתן לראות כי עבור עיוות ערוץ נמוך (a), העיוות הכולל כפונקציה של $\alpha$ מושפע בעיקר מעיוות הקוונטיזציה, ולכן המינימום נמצא קרוב מאד ל-1 = $\alpha$. עבור עיוות ערוץ גבוה (b), משתנה מיקומו של המינימום וניתן לשפר באופן משמעותי את ביצועי המערכת ע"י שימוש בטכניקת הכיווץ.

### 4.3.1 קוונטייזר סקלרי עם כיווץ ליניארי

נסתכל על קוונטייזר סקלרי עם כיווץ ליניארי הפועל בתנאי ערוץ בינארי סימטרי. העיוות הכולל (2.17) עבור מקרה זה הוא:

$$D_{\text{Total}}(\alpha)=\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\pi Q\pi^{T}\right)_{ij}\int_{\alpha\tau_i}^{\alpha\tau_{i+1}}\left(x-\alpha\phi_j\right)^2\cdot f_x(x)\cdot dx \tag{4.2}$$

כאשר $i=0,1,\ldots,N$ , $\tau_i$ הן רמות-ההחלטה של הקוונטייזר, ומוסכם כי $\tau_0=-\infty$ ו- $\tau_N=+\infty$. נניח כי פונקצית צפיפות ההסתברות $f_x(x)$ מקיימת את התנאי: $\left[\log f_x(x)\right]''\le 0$, כלומר הלוגריתם של הפונקציה הוא קעור (log-concave). עבור מקורות מסוג זה, למשל: גאוסי, לפלסי ואחיד, הפונקציה $D_{\text{Total}}(\alpha)$ היא קמורה וקיים מינימום יחיד לפונקציה בנקודה $\alpha=1$ [Fl64], כאשר אין שגיאות בערוץ $q=0$.

בנספח ב' (B.23), מחושבים הערכים העצמיים והוקטורים העצמיים של מטריצת הערוץ $Q$ עבור ערוץ בינארי סימטרי - $Q=V\Lambda V^{T}$. למטריצה $Q$ (עבור $N$ רמות ייצוג ו-$L$ סיביות) קיימים $L+1$ ערכים עצמיים שונים $m=0,1,\ldots,L$ $(1-2q)^m$, כל אחד בריבוי $\binom{L}{m}$.

$$\begin{aligned}
D_{\text{Total}}(\alpha)&=\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\pi V\Lambda V^{T}\pi^{T}\right)_{ij}\int_{\alpha\tau_i}^{\alpha\tau_{i+1}}\left(x-\alpha\phi_j\right)^2\cdot f_x(x)\cdot dx=\\
&\qquad\Leftarrow trace\{AB\}=\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}A_{ij}^{T}B_{ij}\\
&=trace\left\{\pi V\Lambda V^{T}\pi^{T}\Delta(\alpha)\right\}=\\
&\qquad\Leftarrow trace\{AB\}=trace\{BA\}\\
&=trace\left\{\Lambda V^{T}\pi^{T}\Delta(\alpha)\pi V\right\}
\end{aligned} \tag{4.3}$$

כאשר:

$$\left\{\Delta(\alpha)\right\}_{ij}=\int_{\alpha\tau_i}^{\alpha\tau_{i+1}}\left(x-\alpha\phi_j\right)^2\cdot f_x(x)\cdot dx \tag{4.4}$$

הפונקציה $D_{\text{Total}}(\alpha)$ תלויה בהסתברות השגיאה בערוץ $q$. נתאר עובדה זו מפורשות ע"י הסימון $D_{\text{Total}}(\alpha,q)$. הנגזרת השנייה של $D_{\text{Total}}(\alpha,q)$ לפי $\alpha$ היא:

$$\frac{\partial^2}{\partial\alpha^2} D_{\text{Total}}(\alpha,q) = trace\left\{\Lambda V^T \pi^T \Delta''(\alpha)\pi V\right\} \tag{4.5}$$

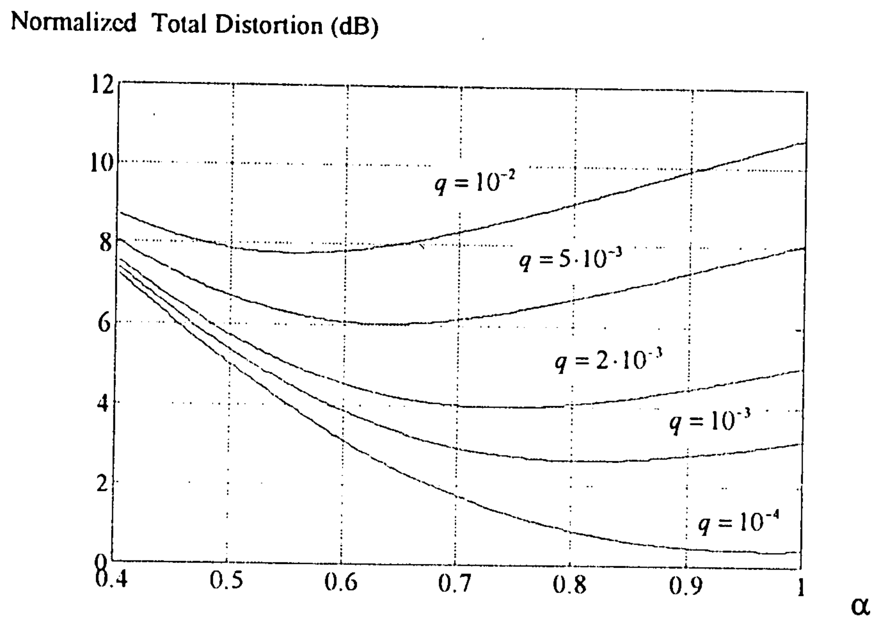
כאשר $\Delta''(\alpha)$ מסמן גזירה של האיברים במטריצה.

כאמור, עבור $q=0$, הנגזרת השנייה לפי $\alpha$ היא חיובית לכל $\alpha$. משיקולי רציפות, הנגזרת השנייה ממשיכה להיות חיובית גם עבור ערכים קטנים של שגיאות בערוץ $q$, ולכן מובטח מינימום יחיד גם עבור ערכי $q$ קטנים. בדוגמאות המספריות המוצגות בהמשך הפרק ניתן לראות כי קיים מינימום יחיד גם בהסתברויות שגיאה של $q=10^{-2}$.

## 4.4 דוגמאות מספריות

### 4.4.1 קוונטייזר סקלרי מותאם לפילוג לפלס

נבחן קוונטייזר סקלרי בן 4 סיביות, לא אחיד, המותאם לפילוג לפלס [Ja84, סעיף 4.4.1]. בציור 4.6 מתואר העיוות הכולל כפונקציה של $\alpha$ עבור קצבי שגיאה שונים בערוץ. הגרף מנורמל לשגיאת הקוונטיזציה.

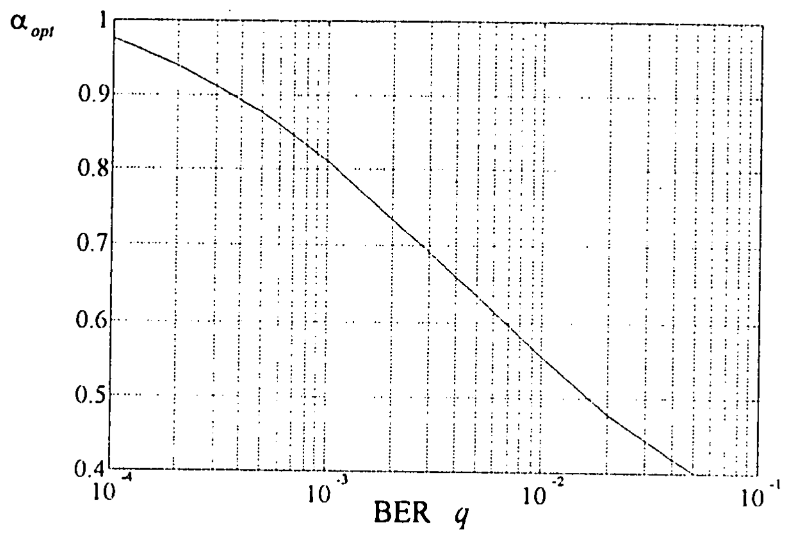



ציור 4.6 - העיוות הכולל בקוונטייזר סקלרי עם כיוון ליניארי בן 4 סיביות המותאם לפילוג לפלס. העיוות מנורמל לעיוות הקוונטיזציה. פרמטר $q$ - קצב השגיאות בערוץ

Fig. 4.6 - Total Distortion for a 4-bit Scaled Laplacian-PDF-optimized Scalar Quantizer. Distortion is normalized to the Quantization Distortion. Parameter $q$ - Channel Bit Error Rate

ניתן להבחין בשיפור בביצועים כתוצאה מכיווץ הקוונטייזר, כאשר השיפור היחסי גדל עם עליית קצב השגיאות בערוץ. בנוסף, הכיווץ נעשה משמעותי יותר עם גידול קצב השגיאות בערוץ. לדוגמה, עבור הסתברות שגיאה של $5\cdot10^{-3}$ ניתן להשיג שיפור של כ-2dB עבור $\alpha_{opt}\approx0.63$ לעומת הקוונטייזר ללא כיווץ (עיוות מנורמל של 6dB במקום 8dB). גרף העיוות הוא שטוח בסביבת המינימום כך שידיעה מדויקת של מיקומו איננה הכרחית. לדוגמה, עבור $q=5\cdot10^{-3}$, שימוש בגורם $\alpha_{opt}\approx0.55$ (האופטימלי ל-$q=10^{-2}$) יתן רווח של כ-1.5dB מתוך כ-2dB שנותן הגורם המתאים.

בציור 4.7 מתואר גורם הכיווץ האופטימלי כפונקציה של קצב השגיאות בערוץ.

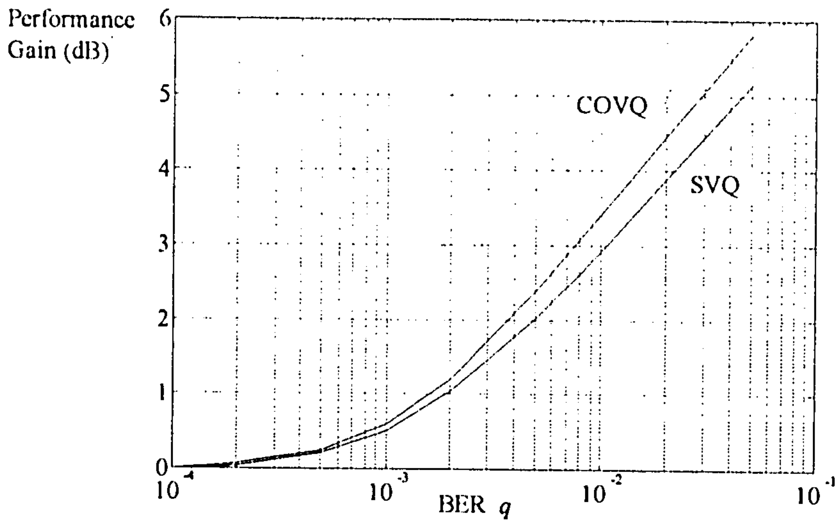



ציור 4.7 - גורם הכיווץ האופטימלי בקוונטייזר סקלרי בן 4 סיביות המותאם לפילוג לפלס

Fig. 4.7 - Scaling Factor for a 4-bit Laplacian-PDF-Optimized Scaled Scalar Quantizer

ניתן לראות כי הכיווץ נעשה משמעותי יותר עם העליה בקצב השגיאות בערוץ. השתנות $\alpha_{opt}$ כפונקציה של הסתברות השגיאה $q$ היא חלקה ובמימוש ניתן לשמור מספר מועט של נקודות בטבלה.

הרווח המתקבל כתוצאה מכיווץ הקוונטייזר כפונקציה של הסתברות השגיאה $q$, בהשוואה לביצועי COVQ, מופיע בציור 4.8.




ציור 4.8 - הרווח בביצועים עבור קוונטייזר סקלרי בן 4 סיביות המותאם לפילוג לפלס

עם כיוון ליניארי (SVQ) בהשוואה לתכן אופטימלי (COVQ)

Fig. 4.8 - Performance gain for a 4-bit Laplacian-PDF-Optimized Scaled Scalar Quantizer (SVQ) Compared with a Channel Optimized Vector Quantizer (COVQ)

ניתן להבחין בציור כי הקוונטייזר המוצע מסוג SVQ, מביא לרווח בביצועים הקרוב לרווח מתכן מותאם לערוץ. למשל עבור $q = 10^{-2}$ משיג ה-SVQ רווח ביצועים של כ-3dB מתוך רווח של כ-3.5dB המושג ע"י ה-COVQ. מצד שני מתאפיין כאמור ה-SVQ בפשטות תכן ובדרישות נמוכות בהרבה בזיכרון הנדרש במימוש.

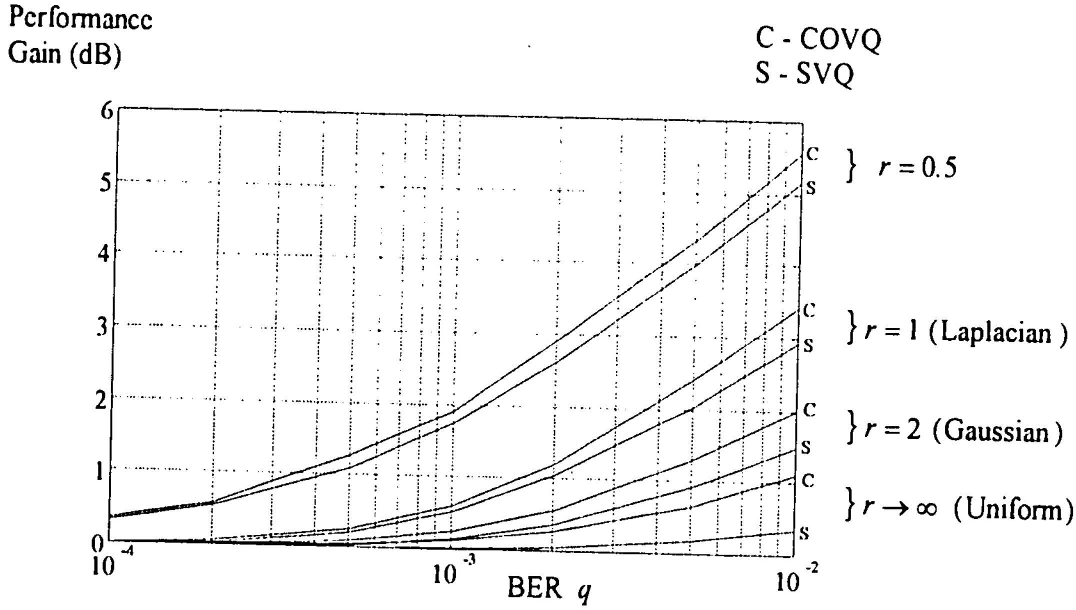
### 4.4.2 קוונטייזר סקלרי מותאם לפילוג גאוסי מוכלל

נתבונן כעת בפילוג גאוסי מוכלל (Generalized Gaussian Distribution) [Cl85, סעיף 4.4.1] עם סטיית תקן השווה לאחד.

$$p(x) = \frac{r}{2}\frac{\Gamma(3/r)^{\frac{1}{2}}}{\Gamma(1/r)^{\frac{3}{2}}} exp\left[-\left(\frac{\Gamma(3/r)}{\Gamma(1/r)}\right)^{\frac{r}{2}} \cdot x^r\right] \qquad (4.6)$$

כאשר $\Gamma(\cdot)$ היא פונקצית גאמה ו- $r$ פרמטר. עבור $r \to 0$ מתקרבת צפיפות הסתברות להיות מרוכזת באפס, עבור $r = 1$ המקור הוא מקור לפלסי, $r = 2$ מקור גאוסי ועבור $r \to \infty$ פילוג ההסתברות שואף לפילוג האחיד.

בציור 4.9 מופיע הרווח בביצועי קוונטייזר סקלרי מותאם לפילוג (עבור ערכי $r$ שונים) כתוצאה מכיווץ ליניארי (SVQ) בהשוואה לקוונטייזר מסוג COVQ.




ציור 4.9 - הרווח בביצועים עבור קוונטייזר סקלרי בן 4 סיביות המותאם לפילוג גאוסי מוכלל עם פרמטר $r$ עם כיווץ ליניארי (SVQ) בהשוואה לקוונטייזר אופטימלי לתנאי הערוץ COVQ

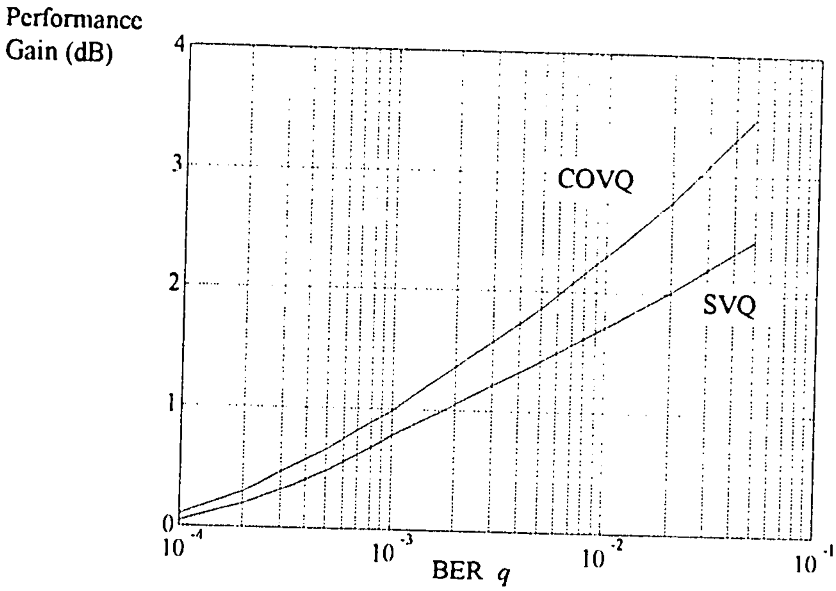
Fig. 4.9 - Performance gain for a 4-bit Generalized-Gaussian-PDF-Optimized Scaled Scalar Quantizer (SVQ) compared with Channel Optimized Vector Quantizer (COVQ). $r$ Parameter of the Generalized-Gaussian PDF

ניתן לראות שככל שצפיפות ההסתברות מתרכזת סביב האפס ( $r$ הולך וקטן) עולה הרווח המושג בכיווץ ליניארי. הסיבה לתופעה זו היא ההבדלים ברעש ההעמסה (Overload). כיווץ הקוונטייזר לצורך הקטנת עיוות הערוץ מצמצם את הכיסוי של הקוונטייזר ומגדיל את רעש ההעמסה. עבור מקורות צפופים, השפעת זנבות הפילוג קטנה יותר, ורעש ההעמסה משתנה באיטיות יחסית עם כיווץ הקוונטייזר. גורם הכיווץ האופטימלי $\alpha_{opt}$ הוא קטן יותר והרווח גדול יותר.

תכן הקוונטייזר האופטימלי (COVQ) מאפשר עיוות שונה במרכז פונקצית ההסתברות ובזנבות, ולכן הרווח בשימוש ב-COVQ יחסית ל-SVQ גדול יותר עבור פונקצית הסתברות עם זנבות משמעותיים יותר, כדוגמת המקור האחיד.

### 4.4.3 קוונטייזר וקטורי מותאם לפילוג גאוסי

נתבון בקוונטייזר וקטורי בעל 256 וקטורים מייצגים (8 סיביות), בקצב של 2 סיביות לדגימה (4 דגימות בוקטור ייצוג), שתואר בציור 3.3. בציור 4.10 מופיע הרווח המתקבל כתוצאה מכיווץ הקוונטייזר כפונקציה של הסתברות השגיאה, $q$, בהשוואה לביצועי COVQ.




ציור 4.10 - הרווח בביצועים עבור קוונטייזר וקטורי בן 8 סיביות (2 סיביות לדגימה) המותאם לפילוג גאוס-מרקוב (מקדם מתאם 0.5) עם כיווץ ליניארי (SVQ)

בהשוואה לתכן אופטימלי (COVQ)

Fig. 4.10 - Performance gain for an 8-bit (2 bit/sample) Scaled Vector Quantizer (SVQ) designed for a Gauss-Markov source (correlation 0.5) Compared with a Channel Optimized Vector Quantizer (COVQ)

גם במקרה זה ניתן לראות כי הקוונטייזר עם כיווץ ליניארי נותן חלק נכבד מן הרווח שהיה ניתן להשיג בתכן אופטימלי מלא. עבור הסתברות שגיאה של $q = 10^{-2}$ נותן ה-SVQ שיפור של כ-1.7dB בביצועים. שיפור נוסף של כ-0.6dB היה מושג ע"י COVQ במחיר גבוה בהרבה במימוש.

## 4.5 סיכום

בפרק זה הצגנו שיטה תת-אופטימלית לשיפור ביצועי קוונטייזר וקטורי תחת שגיאות ערוץ. המוטיבציה לשימוש בטכניקות תת-אופטימליות היא סיבוכיות התכן ומחיר המימוש של התכן האופטימלי. המימוש המוצע מבוסס על קוונטייזר המתוכנן ללא התחשבות בשגיאות ערוץ, ובכיוונון ליניארי של מבנה הקוונטייזר כפונקציה של קצב השגיאות בערוץ. ראינו כי עבור קוונטייזר סקלרי אופטימלי לפילוג המקיים את תנאי ה-log-concave, מובטח מינימום יחיד עבור הסתברויות שגיאה נמוכות בערוץ.

בדוגמאות מספריות ראינו כי ניתן להשיג רווח נאה בביצועים, יחסית לקוונטייזר שלא מתחשב בשגיאות הערוץ. מצד שני, פשטות התכן והמימוש באות על חשבון הרעה בביצועים יחסית לתכן האופטימלי. בדוגמאות שנבחנו ראינו כי הרעה זו היא סבירה ויש למתכנן מקום לשיקול דעת האם לאמץ את השיטה האופטימלית היקרה או לוותר מעט בביצועים ולבחור את השיטה הזולה הרבה יותר.

# פרק 5 - האופטימליות של הקוד-הבינארי-הטבעי עבור קוונטייזר סקלרי אחיד ומקור אחיד (תחת ערוץ בינארי סימטרי)

## 5.1 מבוא

בפרק המבוא תיארנו את החשיבות שבשיבוץ יעיל של מילות קוד לרמות-ייצוג. ראינו גם כי עבור קוונטייזר בעל $N$ רמות ייצוג, מספר השיבוצים האפשריים, $N!$, איננו מאפשר חיפוש ממצה (Exhaustive Search) למציאת השיבוץ האופטימלי, אף עבור קוונטייזר עם מספר רמות לא רב.

בפרק זה נוכיח, בצורה עקיפה, כי הקוד הבינארי הטבעי (Natural Binary Code - NBC) הוא השיבוץ האופטימלי של מילות קוד לרמות הייצוג של קוונטייזר סקלרי אחיד, כאשר מקור האות הוא אחיד והערוץ הוא בינארי סימטרי (Binary Symmetric Channel - BSC). כמו-כן ננסה להסביר מדוע הפתרון "המתבקש" של שיבוץ בקוד Gray איננו אופטימלי.

ההוכחה מופיעה בשלמותה בנספח ב', כאשר בפרק זה נביא את תקציר ההוכחה. תוצאה דומה הוצגה בעבר ע"י [Cr69] ובמקביל ע"י McLaughlin [Mc92]. ההוכחה המוצגת כאן הנה קצרה ופשוטה יותר. בנוסף, הרחבנו את דרך ההוכחה לחסמים המוצגים בפרק 6. חסמים אלו נותנים מידע על מקרים מורכבים יותר: קוונטייזרים וקטוריים וסקלריים כלליים, מקורות כלליים וערוצים חסרי זיכרון נוספים.

## 5.2 השוואה לקוד גריי (Gray)

עבור הערוץ הבינארי הסימטרי, בהסתברויות שגיאה נמוכות, מאורע של שגיאה בסיבית אחת הוא שכיח מאד יחסית למאורעות של שגיאה בשתי סיביות ומעלה. על מנת להקטין את העיוות כתוצאה משגיאות ערוץ, יש לשבץ מילות קוד הנבדלות זו מזו בסיבית בודדת לרמות קרובות ככל האפשר, מבחינת מרחק אוקלידי. קוד Gray [Ja84 סעיף 4.9.1] מאופיין בכך שכל שתי רמות שכנות נבדלות בסיבית אחת בלבד (מאידך, קיימות גם רמות רחוקות אשר נבדלות בסיבית אחת).
קוד זה נראה, לכאורה, מתאים לפתרון לבעיית השיבוץ במקרה זה, אך נראה שאין הדבר כך.

בטבלה 5.1 מתוארים המרחקים הריבועיים (עיוותים) (מנורמלים למרווח בין רמות שכנות) בין רמות הייצוג של קוונטייזר סקלרי אחיד בן $L = 3$ סיביות. כמו-כן רשום הקוד הבינארי הטבעי (NBC - Natural Binary Code) לכל רמת ייצוג, כאשר עיוותים הנגרמים כתוצאה משגיאה בסיבית אחת מסומנים בקו תחתי.

| Code | Level | 000 | 001 | 010 | 011 | 100 | 101 | 110 | 111 |
|---|---|---|---|---|---|---|---|---|---|
| | Level | −4 | −3 | −2 | −1 | 0 | 1 | 2 | 3 |
| 000 | −4 | 0 | 1 | 4 | 9 | 16 | 25 | 36 | 49 |
| 001 | −3 | 1 | 0 | 1 | 4 | 9 | 16 | 25 | 36 |
| 010 | −2 | 4 | 1 | 0 | 1 | 4 | 9 | 16 | 25 |
| 011 | −1 | 9 | 4 | 1 | 0 | 1 | 4 | 9 | 16 |
| 100 | 0 | 16 | 9 | 4 | 1 | 0 | 1 | 4 | 9 |
| 101 | 1 | 25 | 16 | 9 | 4 | 1 | 0 | 1 | 4 |
| 110 | 2 | 36 | 25 | 16 | 9 | 4 | 1 | 0 | 1 |
| 111 | 3 | 49 | 36 | 25 | 16 | 9 | 4 | 1 | 0 |

טבלה 5.1 - מרחקים ריבועיים מנורמלים בין רמות הייצוג בקוונטייזר סקלרי אחיד.
לכל רמת ייצוג מסומן השיבוץ בקוד הבינארי הטבעי.
עיוותים הנגרמים כתוצאה משגיאה בסיבית אחת מצוינים ע"י קו-תחתי.

Table 5.1 - Normalized squared distances among representing-levels of a uniform scalar quantizer. Each level is assigned with its Natural-Binary-Code codeword. Distortions due to a single-bit errors are underlined.

העיוות הממוצע כתוצאה משגיאה (של סיבית אחת) במקרה זה הוא:
$(8\cdot 1+8\cdot 4+8\cdot 16)/24 = 7$.

נבחן קוד Gray בן $L$ סיביות ($N = 2^L$ רמות ייצוג). עבור כל רמת ייצוג (פרט לקצוות) קיימת סיבית אחת, אשר שגיאה בה תגרור שחזור של רמת ייצוג שכנה כלפי מעלה. קיימת גם סיבית אחרת, אשר שגיאה בה תגרור שחזור של רמת ייצוג שכנה כלפי למטה. שגיאה בשאר $L$-2 הסיביות במילת הקוד יגרמו לשחזור של רמות ייצוג רחוקות יותר. לדוגמה, הקודים של רמות הייצוג בקצוות (000...0 ו- 100...0) נבדלים ביניהם בסיבית אחת.

בטבלה 5.2 רשום קוד Gray לכל רמת ייצוג. עיוותים הנגרמים כתוצאה משגיאה בסיבית אחת מסומנים בקו תחתי. יש לשים לב כי במקרה זה, עבור כל רמת ייצוג, שגיאה בשתיים מתוך שלוש הסיביות יגרמו לשחזור רמת-ייצוג שכנה.

| Code | | 000 | 001 | 011 | 010 | 110 | 111 | 101 | 100 |
|---|---|---|---|---|---|---|---|---|---|
| | Level | −4 | −3 | −2 | −1 | 0 | 1 | 2 | 3 |
| 000 | −4 | 0 | $\underline{1}$ | 4 | $\underline{9}$ | 16 | 25 | 36 | $\underline{49}$ |
| 001 | −3 | $\underline{1}$ | 0 | $\underline{1}$ | 4 | 9 | 16 | $\underline{25}$ | 36 |
| 011 | −2 | 4 | $\underline{1}$ | 0 | $\underline{1}$ | 4 | $\underline{9}$ | 16 | 25 |
| 010 | −1 | $\underline{9}$ | 4 | $\underline{1}$ | 0 | $\underline{1}$ | 4 | 9 | 16 |
| 110 | 0 | 16 | 9 | 4 | $\underline{1}$ | 0 | $\underline{1}$ | 4 | $\underline{9}$ |
| 111 | 1 | 25 | 16 | $\underline{9}$ | 4 | $\underline{1}$ | 0 | $\underline{1}$ | 4 |
| 101 | 2 | 36 | $\underline{25}$ | 16 | 9 | 4 | $\underline{1}$ | 0 | $\underline{1}$ |
| 100 | 3 | $\underline{49}$ | 36 | 25 | 16 | $\underline{9}$ | 4 | $\underline{1}$ | 0 |

טבלה 5.2 - מרחקים ריבועיים מנורמלים בין רמות הייצוג בקוונטייזר סקלרי אחיד.
לכל רמת ייצוג מסומן השיבוץ בקוד Gray.
עיוותים הנגרמים כתוצאה משגיאה בסיבית אחת מצוינים ע"י קו-תחתי.

Table 5.2 - Normalized squared distances among representing-levels of a uniform scalar quantizer. Each level is assigned with its Gray-code codeword. Distortions due to a single-bit errors are underlined.

העיוות הממוצע כתוצאה משגיאה (של סיביית אחת) במקרה זה הוא:

$(14\cdot 1+6\cdot 9+2\cdot 25+2\cdot 49)/24=9$.

עיקר העיוות נגרם כתוצאה מכך שבקוד Gray רמות הייצוג הרחוקות (-4,+3) ו-(-3,+2) נבדלות בסיבית אחת בלבד.

בדוגמה זו ניתן לראות כי עיוות גבוה יותר נגרם כתוצאה מקוד Gray בהשוואה לקוד הבינארי הטבעי. כעת נפנה להוכחה כי **השיבוץ האופטימלי הוא הקוד הבינארי הטבעי**. תוצאה זו תהיה נכונה לכל הסתברות שגיאה $q$ בערוץ בתחום $0\le q<\frac{1}{2}$.

## 5.3 תקציר ההוכחה

נתאר כעת בקצרה את ההוכחה לשיבוץ האופטימלי. כאמור, ההוכחה במלואה מופיעה **בנספח ב'**. ההוכחה מבוססת על החלפת הבעיה הקומבינטורית בבעיית אופטימיזציה רציפה. כפי שניווכח לדעת, במקרה זה (קוונטייזר סקלרי אחיד, מקור אות אחיד וערוץ בינארי סימטרי) האופטימום בשתי הבעיות מתלכד.

שלבי ההוכחה הם:

1. תיאור הבעיה ודרך למציאת חסמים ע"י מעבר לבעיית אופטימיזציה רציפה.
2. פתרון הבעיה הרציפה ע"י כופלי לגרנג' והגדרת החסמים.
3. הוכחת התלכדות החסם התחתון עם ביצועי הקוד הבינארי הטבעי.

### 5.3.1 תיאור הבעיה ומעבר לרצף

נתבונן בקוונטייזר סקלרי בן $L$ סיביות ו-$N = 2^L$ רמות ייצוג. רמות הייצוג מתוארות בציור 3.7. נתאר את רמות הייצוג בוקטור עמודה, כמו ב-(3.11):

$$\underline{\rho} = \begin{bmatrix} \rho_0 \\ \rho_1 \\ \vdots \\ \rho_i \\ \vdots \\ \rho_{N-1} \end{bmatrix} = \frac{2}{N} \cdot \begin{bmatrix} -N/2+1/2 \\ 1-N/2+1/2 \\ \vdots \\ i-N/2+1/2 \\ \vdots \\ N/2-1+1/2 \end{bmatrix} \qquad (5.1)$$

ביטוי כללי לעיוות הערוץ הוא, עפ"י (2.15):

$$D_c = trace\left\{P\pi Q_N \pi^T D\right\} \qquad (5.2)$$

במקרה הנוכחי ניתן להציב את הגדלים הבאים:

$$P = \frac{1}{N} I$$
$$\{D\}_{ij} = \left(\underline{\rho}_i - \underline{\rho}_j\right)^2$$
$$\{Q_N\}_{ij} = q^{H(i,j)}(1-q)^{L-H(i,j)} \qquad (5.3)$$

כאשר $I$ היא מטריצת היחידה, $L = log_2 N$ - מספר הסיביות, $q$ - היא הסתברות השגיאה בערוץ ו-$H(i,j)$ הוא מרחק Hamming בין הייצוג הבינארי של $i$ ו-$j$.

הצבת הגדלים ב-(5.2) ושימוש בזהויות אלגבריות (ראה B.3) מביא את ביטוי עיוות הערוץ לצורה הבאה:

$$D_c = \frac{2}{N}\left(\underline{z}^T \underline{z} - \underline{z}^T Q_N \underline{z}\right) \tag{5.4}$$

כאשר הוקטור $\underline{z} = \pi\underline{\rho}$ הוא פרמוטציה של וקטור רמות הייצוג.

על מנת למצוא חסם עליון וחסם תחתון לעיוות הערוץ **מעל כל שיבוצי מילות הקוד האפשריים,** ניתן לאפשר לאיברי $\underline{z}$ להיות רציפים, תוך שמירת שני המומנטים הראשונים. בכך נעבור מבעיית אופטימיזציה דיסקרטית מעל $N!$ אפשרויות לבעיית אופטימיזציה רציפה. יש לשים לב כי עבור כל הפרמוטציות האפשריות מתקיים:

א. סכום איברי $\underline{z}$ קבוע ושווה לאפס.

ב. הנורמה של הוקטור קבועה: $\underline{z}^T \underline{z} = \underline{\rho}^T \pi^T \cdot \pi\underline{\rho} = \underline{\rho}^T \underline{\rho} = \frac{(N+1)(N-1)}{3N} = k^2$.

החסמים ימצאו ע"י פתרון בעיית האופטימיזציה הבאה:

$$\begin{aligned} \underset{\underline{z}}{\min}/\underset{\underline{z}}{\max} \quad & f(\underline{z}) = \underline{z}^T \underline{z} - \underline{z}^T Q_N \underline{z} \\ \text{s.t.} \quad & \underline{1}^{\mathrm{T}} \underline{z} = 0 \\ & \underline{z}^T \underline{z} = \underline{\rho}^T \underline{\rho} = k^2 \end{aligned} \tag{5.5}$$

כאשר $\underline{1}$ הוא וקטור $N$-ממדי שכל איבריו הם אחד.

האיבר הראשון בפונקצית המטרה $\underline{z}^T \underline{z}$ הוא, כאמור, קבוע. ניתן להוציא אותו מן הבעיה ונקבל:

$$\begin{aligned} \underset{\underline{z}}{\max}/\underset{\underline{z}}{\min} \quad & f(\underline{z}) = \underline{z}^T Q_N \underline{z} \\ \text{s.t.} \quad & \underline{1}^{\mathrm{T}} \underline{z} = 0 \\ & \underline{z}^T \underline{z} = \underline{\rho}^T \underline{\rho} = k^2 \end{aligned} \tag{5.6}$$

## 5.3.2 פתרון בעיית האופטימיזציה באמצעות כופלי לגרנג'

לפני שניגש לפתרון נשתמש בתכונת הסימטריה של מטריצת הערוץ $Q_N$. למטריצה יש לכסון יוניטרי:

$$Q_N = V \cdot \Lambda \cdot V^T \qquad V \cdot V^T = I \tag{5.7}$$

מכיוון שמטריצת הערוץ $Q_N$ מייצגת הסתברויות, סכום כל השורות שלה הוא אחד כלומר $Q_N \underline{1} = \underline{1}$. נניח לכן כי הערך העצמי הראשון של המטריצה $\lambda_0 = 1$ והוקטור העצמי המתאים (העמודה הראשונה של $V$) $\underline{v}_0 = \frac{1}{\sqrt{N}}\underline{1}$.

נפתור כעת את (5.6) ע"י כופלי לגרנג' $\mu_1, \mu_2$:

$$L(\underline{z}, \mu_1, \mu_2) = \underline{z}^T Q_N \underline{z} + \mu_1\left(\underline{z}^T \underline{z} - k^2\right) + \mu_2\left(\underline{1}^T \underline{z}\right) \tag{5.8}$$

פתרון הבעיה מוביל ל-$N$ המשוואות הבאות, המופיעות בנספח ב' (B.17):

$$\left[\lambda_i - \left(\underline{\xi}^T \Lambda \underline{\xi}\right)\right] \cdot \xi_i = 0, \quad i = 0,1,2,\ldots,N-1 \tag{5.9}$$

כאשר הוקטור $\underline{\xi}$ מוגדר ע"י $\underline{\xi} = \frac{1}{k} V^T \underline{z}$. המשוואה הראשונה, זו המתאימה ל-$\lambda_0 = 1$, נכונה תמיד מכיוון שהאיבר הראשון בוקטור שווה לאפס (B.12): $\xi_0 = 0$.

נניח כעת כי לערך עצמי מסוים של $Q_N$ יש ריבוי $m$: $\lambda_l = \lambda_{l+1} = \cdots = \lambda_{l+m-1} = \lambda$. פתרון המתאים לערך עצמי זה הוא כדלקמן:

איברי $\underline{\xi}$ שאינם מתאימים לערך העצמי $\lambda$ שווים לאפס: $\xi_i = 0 \quad i \neq l, l+1, \ldots, l+m-1$

איברי $\underline{\xi}$ המתאימים לערך העצמי $\lambda$ נקבעים כך שיקיימו את הדרישה: $\underline{\xi}^T \underline{\xi} = 1$ (B.12a). קיום הדרישה מבטיח כי $\underline{\xi}^T \Lambda \underline{\xi} = \lambda$.

הוקטור המקורי $\underline{z} = k \cdot \underline{w} = k \cdot V \underline{\xi}$ מורכב מקומבינציה ליניארית של עמודות $V$ ולכן של וקטורים עצמיים של $Q_N$, אלו המתאימים לערך העצמי $\lambda$. הצבת $\underline{z}$ בפונקצית המטרה של הבעיה המקורית (5.5) תיתן:

$$\begin{aligned} f(\underline{z}) &= k^2 - k^2 \underline{\xi}^T V^T \cdot V \Lambda V^T \cdot V \underline{\xi} = \\ &= k^2\left(1 - \underline{\xi}^T \Lambda \underline{\xi}\right) = k^2 (1 - \lambda) \end{aligned} \tag{5.10}$$

המקסימום יתקבל עבור הערך העצמי $\lambda$ הקטן ביותר (פרט ל-$\lambda_0 = 1$). המינימום יתקבל עבור הערך העצמי $\lambda$ הגדול ביותר (פרט ל-$\lambda_0 = 1$).

בנספח ב׳ מתואר חישוב הערכים העצמיים והוקטורים העצמיים של המטריצה $Q_N$ באופן רקורסיבי (עבור הסתברויות שגיאה בערוץ המקיימות $q < 0.5$). עבור קוונטייזר בן סיבית אחת, $L$=1, הוקטורים העצמיים והערכים העצמיים המתאימים הם:

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}, 1 \quad ; \quad \begin{bmatrix} 1 \\ -1 \end{bmatrix}, 1-2q \tag{5.11}$$

נניח כי $\underline{v}$ הוא וקטור עצמי של המטריצה $Q_N$ ($L$ סיביות) המתאים לערך עצמי α. אזי ניתן לבנות ממנו (ראה B.22) שני זוגות של וקטורים עצמיים וערכים עצמיים עבור המטריצה $Q_{2N}$ ($L$+1 סיביות). וקטור עצמי אחד מתקבל ע״י ביצוע שכפול והוקטור העצמי השני ע״י שכפול שלילי:

$$\begin{bmatrix} \underline{v} \\ \underline{v} \end{bmatrix}, \alpha \quad ; \quad \begin{bmatrix} \underline{v} \\ -\underline{v} \end{bmatrix}, (1-2q)\alpha \tag{5.12}$$

מהסתכלות על תהליך הבניה הרקורסיבית של מטריצת הלכסון (B.23), ניתן להבחין כי היא מטריצת Walsh-Hadamard המופיעה בשטחי עיבוד תמונה [C185 סעיף 7.3].

כללית יש למטריצה $Q_N$ ($L$ סיביות) $L+1$ ערכים עצמיים שונים $m=0,1,\ldots,L$ $(1-2q)^m$, כל אחד בריבוי $\binom{L}{m}$.

נציב את הערכים העצמיים ב-(5.10) כאשר הערך העצמי הגדול ביותר פרט ל- 1 הוא $1-2q$, והערך העצמי הקטן ביותר הוא $(1-2q)^L$.

בצורה זו נקבל את החסמים על העיוות כתוצאה משגיאות ערוץ:

$$\frac{2(N+1)(N-1)}{3N^2}\cdot 2q \le D_c \le \frac{2(N+1)(N-1)}{3N^2}\left[1-(1-2q)^L\right] \tag{5.13}$$

היחס בין החסמים עבור הסתברויות שגיאה קטנות בערוץ הוא:

$$\lim_{q\to 0}\frac{\left[1-(1-2q)^L\right]}{2q} = L \tag{5.14}$$

כלומר, <u>היחס בין החסמים גדל ליניארית עם מספר הסיביות</u>.

בפרק הבא, בסעיף 6.3.1, קיים חישוב מספרי של החסמים הללו עבור קוונטייזר אחיד בן $L=4$ סיביות. כמו-כן נעשית שם השוואה בין החסמים לבין תוצאות 10,000 שיבוצים אקראיים (מתוך כ-$16!\approx 2\cdot 10^{13}$). המרחק בין החסמים הוא $10\cdot log_{10}4=6dB$. החסם התחתון מושג ע"י השיבוץ הבינארי הטבעי והחסם העליון רחוק אך כ-0.5dB מן השיבוץ הגרוע ביותר שנתגלה. המסקנה היא שבמקרה זה ניתן לקבל רווח של כ-$5.5-6dB$ ע"י בחירת שיבוץ טוב לעומת בחירה גרועה.

### 5.3.3 הוכחת התלכדות ביצועי הקוד הבינארי הטבעי עם החסם התחתון

בסעיף הקודם הוכחנו כי המינימום לבעיית האופטימיזציה הרציפה מתקבל עבור וקטורים עצמיים של המטריצה $Q_N$ המתאימים לערך העצמי $1-2q$. נראה כעת כי הוקטור $\underline{\rho}=\underline{z}$, המתאים לקוד הבינארי הטבעי, הוא אכן וקטור עצמי כזה.
וקטור עצמי המתאים לערך העצמי $1-2q$ נבנה ע"י $L-1$ שכפולים ושכפול-שלילי אחד (5.12). כאשר השכפול הראשון הוא השכפול שלילי מתקבל הוקטור:

$$\underline{v}_1=\begin{bmatrix}1 & -1 & 1 & -1 & 1 & -1 & \cdots & \cdots & 1 & -1\end{bmatrix}^T \tag{5.15}$$

איברי הוקטור הם $(-1)^i,\ i=0,1,\cdots,N-1$.
כאשר השכפול השני הוא השלילי מתקבל הוקטור:

$$\underline{v}_2=\begin{bmatrix}1 & 1 & -1 & -1 & 1 & 1 & \cdots & \cdots & -1 & -1\end{bmatrix}^T \tag{5.16}$$

איברי הוקטור הם $(-1)^{\lfloor i/2\rfloor},\ i=0,1,\cdots,N-1$ כאשר $\lfloor i/2\rfloor$ מתאר חילוק שלמים ללא שארית.
הוקטור העצמי האחרון המתאים לערך העצמי $1-2q$ נבנה כאשר השכפול האחרון הוא שלילי. מתקבל הוקטור:

$$\underline{v}_L=\begin{bmatrix}1 & 1 & \cdots & 1 & 1 & -1 & -1 & \cdots & -1 & -1\end{bmatrix}^T \tag{5.17}$$

קיים דמיון בין שיטת הספירה הבינארית לבין הוקטורים העצמיים. איברי וקטור $\underline{v}_1$ מתנהגים כמו הספרה הבינארית הפחות משמעותית (LSB - Least Significant Bit) בספירה בינארית. איברי הוקטור האחרון, $\underline{v}_L$ מתנהגים כמו הספרה הבינארית המשמעותית ביותר (MSB - Most Significant Bit).

נבנה קומבינציה ליניארית של הוקטורים אשר תשווה לוקטור רמות הייצוג $\underline{\rho}$:

$$\underline{z} = \frac{1}{N}\sum_{i=1}^{L} 2^{i-1} \cdot \underline{v}_i = \frac{1}{N}\begin{bmatrix}-1 & 1 & -1 & 1 & -1 & 1 & \cdots & \cdots & -1 & 1\end{bmatrix}^T +$$
$$+\frac{2}{N}\begin{bmatrix}-1 & -1 & 1 & 1 & -1 & -1 & \cdots & \cdots & 1 & 1\end{bmatrix}^T +$$
$$\vdots$$
$$+\frac{1}{2}\begin{bmatrix}-1 & -1 & \ldots & -1 & -1 & 1 & 1 & \cdots & 1 & 1\end{bmatrix}^T = \underline{\rho} \tag{5.18}$$

מכיוון ש-$\underline{\rho}$ הוא וקטור עצמי המתאים לערך העצמי $1-2q$, הוא מהווה פתרון אופטימלי לבעיית האופטימיזציה הרציפה. וקטור זה מתאים לשיבוץ חוקי של מילות קוד - הקוד הבינארי הטבעי. מכאן שהקוד הבינארי הטבעי הוא השיבוץ האופטימלי.

מעניין לציין כי <u>השיבוץ הבינארי הטבעי איננו השיבוץ האופטימלי היחיד</u>. בנספח ב' (B.31) מוצגת בנייה של $2^L \cdot L!$ שיבוצים שונים בעלי ביצועים זהים לקוד הבינארי הטבעי.

## 5.4 סיכום

בפרק זה הוכחנו כי השיבוץ הבינארי הטבעי (NBC) הוא שיבוץ אופטימלי עבור קוונטייזר סקלרי אחיד, מקור אחיד וערוץ בינארי סימטרי. דוגמה מספרית מפורטת למקרה הזה ניתן למצוא בהמשך בסעיף 6.3.1. בסעיף 6.3.2 נראה עוד כי השיבוץ הבינארי הטבעי אופטימלי גם עבור בעיה אחרת, דומה, אך השונה מבחינת תנאי הערוץ.

ההוכחה כי ה-NBC הוא השיבוץ האופטימלי במקרה הנדון, נעשתה בטכניקה של החלפת בעיית האופטימיזציה הדיסקרטית בבעיה רציפה. שימוש בטכניקה דומה יאפשר לנו למצוא חסמים למקרים כלליים יותר בפרק הבא.

# פרק 6 - חסמים על עיוות הערוץ מעל כל שיבוצי הקודים האפשריים (ערוץ סימטרי חסר זיכרון)

## 6.1 מבוא

בפרקים הקודמים עמדנו על מספרם העצום של השיבוצים האפשריים של מילות קוד לוקטורי-ייצוג בקוונטייזרים וקטוריים. ראינו בפרק 5 כי במקרה המיוחד של קוונטייזר סקלרי אחיד, מקור אחיד וערוץ בינארי סימטרי, ניתן להצביע על השיבוץ האופטימלי, שהוא הקוד הבינארי הטבעי, וניתן לקבל חסם עליון על ביצועי השיבוץ הגרוע ביותר (5.13).

בפרק זה יתוארו חסמים על עיוות הערוץ מעל כל $N!$ השיבוצים האפשריים של מילות קוד בקוונטייזרים וקטוריים כלשהם. החסמים ניתנים ליישום עבור ערוצים חסרי-זיכרון סימטריים. אין כל הגבלה על מבנה הקוונטייזרים, מידת המרחק ועל מקורות-האות. החסמים יכולים לתת למתכנן מידע כללי על עיוות הערוץ הצפוי במערכת לפני ההחלטה על שיבוץ מסוים. ניתן לעמוד על מרווח-הביצועים שניתן לקבל תוך שינוי השיבוצים ולבחון שיבוץ קיים באמצעות מיקומו בין החסמים.

פיתוח החסמים מופיע בשלמותו **בנספח ג'**, בפרק זה נציג את תקציר הפיתוח בלבד. טכניקת החסימה דומה לטכניקה שהוצגה בפרק 5 בכך שאנו מחליפים את בעיית האופטימיזציה הדיסקרטית בבעיית אופטימיזציה רציפה הניתנת לפתרון בשיטות "קלאסיות". עבור המקרה של קוונטייזר סקלרי אחיד ומקור אחיד מתלכדים החסמים עם התוצאות שהוצגו בפרק 5. בסופו של הפרק ניתנות דוגמאות מספריות עבור קוונטייזרים שונים וערוצים שונים.

## 6.2 תקציר פיתוח החסמים

בדומה לפרק 5, טכניקת החסימה מבוססת על החלפת בעיית האופטימיזציה הקומבינטורית בבעיית אופטימיזציה רציפה. שלבי הפיתוח הם:

1. הצגת עיוות הערוץ ע"י מטריצה מרחקים משוקללת $\hat{D}$. $\hat{D}$ מכילה את המידע על המרחקים בין וקטורי הייצוג ואת הסתברויות ההופעה שלהן במקור.
2. החלפת המטריצה $\hat{D}$ בבעיה למטריצה סימטרית אחרת $\tilde{D}$ בעלת התכונה הבאה:

המטריצה $\tilde{D}$ נבנית כך שהווקטור: $\underline{1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$ הוא וקטור עצמי שלה. תכונה זו נדרשת לשלב האחרון של הפיתוח.

3. הרחבת הבעיה הקומבינטורית לבעיה רציפה.
4. פתרון הבעיה הרציפה תוך שימוש בוקטורים העצמיים של המטריצה $\tilde{D}$ ומציאת החסמים.

### 6.2.1 תיאור הבעיה

לפי (2.15), ביטוי כללי לעיוות הערוץ ניתן ע"י:

$$D_c = trace\left\{P\pi Q_N \pi^T D\right\} \tag{6.1}$$

אנו מניחים כי מטריצת הערוץ $Q_N$ הינה סימטרית. מכיוון שהיא מייצגת הסתברויות, סכום כל שורה שווה לאחד ולכן $Q_N \cdot \underline{1} = \underline{1}$. אין הנחות נוספות על מטריצת ההסתברויות $P$ (אלכסונית) ועל מטריצת המרחקים בין הווקטורים המיצגים $D$. השיבוץ מתואר ע"י מטריצת הפרמוטציה $\pi$.

תוך שימוש בזהויות אלגבריות (C.1) ניתן לאחד את המטריצות $P$ ו-$D$ למטריצה סימטרית אחת $\hat{D}$ - כך שמתקיים:

$$D_c = \frac{1}{2} trace\left\{Q_N \pi^T \hat{D} \pi\right\} \tag{6.2}$$

כאשר $\hat{D} = DP + P^T D^T$ היא מטריצת המרחקים המשוקללת.

### 6.2.2 שינוי במטריצת המרחקים המשוקללת

**הקדמה**

בסעיף 6.2.4 נפתור בעיית אופטימיזציה עם אילוצים. אם $\underline{1}$ הוא וקטור עצמי של מטריצת המרחקים המשוקללת, פתרון הבעיה נעשה פשוט יותר והפתרון הוא פונקציה של הערכים העצמיים של מטריצת המרחקים המשוקללת. מסיבה זו, כשלב ראשון, נשנה את בעיית האופטימיזציה כך שמטריצה $\tilde{D}$ תחליף את $\hat{D}$. הוקטור $\underline{1}$ יהיה וקטור עצמי של המטריצה החדשה $\tilde{D}$.

בנוסף (C.2) מוגדרת מטריצה $C_i$ בגודל $N \times N$ המכילה "1"-ים בעמודה ה-$i$, שאר המטריצה מורכבת מאפסים:

$$C_i = \begin{bmatrix} 0 & & 0 & 1 & 0 & & & 0 \\ 0 & & 0 & 1 & 0 & & & 0 \\ 0 & & 0 & 1 & 0 & & & 0 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & & 0 & 1 & 0 & & & 0 \\ 0 & & 0 & 1 & 0 & & & 0 \\ 0 & & 0 & 1 & 0 & & & 0 \end{bmatrix}$$
$$\uparrow \; i-\text{th column} \tag{6.3}$$

מכיוון ש-$Q_N \cdot \underline{0} = \underline{0}$ וכן $Q_N \cdot \underline{1} = \underline{1}$, מטריצה זו מקיימת את התכונה $Q_N \cdot C_i = C_i$.

בנוסף (C.6) ניתן לראות כי תוספת של מטריצה במבנה צלב: $\alpha\left(C_i + C_i^T\right)$, כאשר $\alpha$ סקלר כלשהו, ל-(6.2) גורמת לתוספת קבוע אשר איננו תלוי בשיבוץ, ולכן איננו משנה את מהות הבעיה:

$$\frac{1}{2}trace\left\{Q_N\pi^T\left(\hat{D}+\alpha C_i^T+\alpha C_i\right)\pi\right\}= \\ =\frac{1}{2}trace\left\{Q_N\pi^T\hat{D}\pi\right\}+\alpha \tag{6.4}$$

על מנת שהווקטור $\underline{1}$ יהיה וקטור עצמי של המטריצה $\tilde{D}$, על סכום האיברים בשורה להיות שווה בכל השורות.

תוספת מטריצה במבנה צלב $\alpha\left(C_i + C_i^T\right)$ מוסיפה את הסקלר $\alpha$ לסכום כל השורות, פרט לשורה ה-$i$ אשר מקבלת תוספת של $\alpha(N+1)$ לסכום איבריה. סכום השורה ה-$i$ מקבל בכך "יתרון" של $N\alpha$ לעומת שאר השורות. תוספת המטריצה במבנה הצלב שומרת על הסימטריה של $\hat{D}$.

בעזרת האלגוריתם הבא אנו מוסיפים למטריצה $\hat{D}$ מטריצות במבנה צלב אשר יביאו את הסכום בכל שורה ושורה להיות שווה לסכום הגדול ביותר. את סכום הסקלרים המתווספים יש לשמור ולדאוג להחסירו מן הביטוי לעיוות.

**<u>אלגוריתם להשוואת סכומי האיברים בשורה עבור כל השורות במטריצת המרחקים המשוקללת</u>**

<u>אתחול</u>: מטריצת המרחקים המשוקללת $\tilde{D} \leftarrow \hat{D} = DP + P^TD^T$

סכום הסקלרים המתווספים $S \leftarrow 0$.

<u>צעד 1</u>: חשב את הסכום $S_i$ של כל השורות במטריצה $\tilde{D}$. $S_i = \sum_{j=0}^{N-1}\tilde{D}_{ij}\,,\quad i=0,1,\ldots,N-1$.

<u>צעד 2</u>: חפש את הסכום המקסימלי וסמן את השורה עם סכום זה ב-$k$. $S_k = \max_{i=0,1,\ldots,N-1} S_i$.

<u>צעד 3</u>: עבור כל שורה $i \neq k$:

הוסף מטריצה במבנה צלב: $\tilde{D} \leftarrow \tilde{D} + \frac{1}{N}(S_k - S_i)\left(C_i + C_i^T\right)$.

עדכן את סכום הסקלרים: $S \leftarrow S + \frac{1}{N}(S_k - S_i)$.

בסיום האלגוריתם המטריצה הסימטרית $\tilde{D}$ מקיימת את התכונה הדרושה $\tilde{D}\cdot\underline{1} = \omega_0\cdot\underline{1}$, כאשר $\omega_0$ הוא הערך העצמי של $\tilde{D}$ המתאים לוקטור העצמי $\underline{1}$, ו-$S$ הוא סכום הסקלרים שהתווספו.

עיוות הערוץ כפונקציה של $\tilde{D}$ הוא:

$$D_c = \frac{1}{2} trace\{Q_N \pi^T \tilde{D} \pi\} - S \tag{6.5}$$

כאשר $S$ הוא סכום הסקלרים שחושב בצעד 3 של האלגוריתם.

### 6.2.3 מעבר לבעיית אופטימיזציה רציפה

מטריצת הערוץ $Q_N$ הנה חיובית וסימטרית ולכן ניתן ללכסן אותה לכסון יוניטרי.

$$Q_N = V \cdot \Lambda \cdot V^T \qquad V \cdot V^T = I \tag{6.6}$$

מכיוון ש-$Q_N \cdot \underline{1} = \underline{1}$, נוכל לבחור שהערך העצמי הראשון יהיה $\lambda_0 = 1$ והוקטור העצמי (עמודה ראשונה של המטריצה $V$) יהיה $\frac{1}{\sqrt{N}}\underline{1}$. עיוות הערוץ מבוטא כעת:

$$\begin{aligned} D_c &= \frac{1}{2} trace\{V \Lambda V^T \pi^T \tilde{D} \pi\} - S = \\ &= \frac{1}{2} trace\{\pi V \Lambda V^T \pi^T \tilde{D}\} - S = \end{aligned} \tag{6.7}$$

המטריצה $\pi V$ היא מטריצה יוניטרית לכל פרמוטציה $\pi$, מכיוון ש-$(\pi V)(\pi V)^T = \pi V V^T \pi^T = \pi \pi^T = I$. העמודה הראשונה של המטריצה, $\frac{1}{\sqrt{N}}\underline{1}$, איננה משתנה תחת פרמוטציה. על מנת לקבל את החסמים, נרחיב את המשפחה הדיסקרטית של מטריצות המתקבלות מ-$V$ ע"י פרמוטציה של שורות, למשפחת כלל המטריצות היוניטריות כאשר העמודה הראשונה שלהן זהה למטריצה המקורית. נחפש את המינימום והמקסימום של הביטוי המורחב לעיוות הערוץ ונקבל חסמים על הביצועים מעל כל הפרמוטציות האפשריות. בפתרון בעיית האופטימיזציה ניתן להתעלם מן הקבוע $S$.

יש לכן לפתור את הבעיה הבאה:

$$\min_U / \max_U \left( trace\{\Lambda U^T \tilde{D} U\} \right) \tag{6.8}$$

כאשר $U$ היא מטריצה יוניטרית, $UU^T = I$, שהעמודה הראשונה שלה היא $\underline{u}_0 = \frac{1}{\sqrt{N}}\underline{1}$.

### 6.2.4 פתרון בעיית אופטימיזציה הרציפה

בנספח ג' (C.11) ניתנת הוכחה כי המינימום והמקסימום ל-(6.8) מתקיימים כאשר המטריצה $U$ מורכבת מווקטורים עצמיים של $\tilde{D}$. נסמן את הערכים העצמיים של $\tilde{D}$ ע"י $\omega_i$, $i = 0,1,\dots,N-1$ ואת הוקטורים העצמיים המתאימים ע"י $\underline{z}_i$, $i = 0,1,\dots,N-1$. כאן עומדת לרשותנו העובדה שהמטריצה $\tilde{D}$ נבנתה כך שהוקטור $\underline{z}_0 = \frac{1}{\sqrt{N}}\underline{1}$ הוא וקטור עצמי שלה והאילוץ על $U$ מתקיים.

ערך פונקצית המטרה ב-(6.8) הוא:

$$trace\{\Lambda U^T \tilde{D} U\} = \sum_{i=0}^{N-1} \lambda_i \underline{z}_i^T \tilde{D} \underline{z}_i = \sum_{i=0}^{N-1} \lambda_i \omega_i \underline{z}_i^T \underline{z}_i = \\ = \sum_{i=0}^{N-1} \lambda_i \omega_i \tag{6.9}$$

עדין נותרה בידינו דרגת החופש של בחירת סדר הערכים העצמיים. פרט ל-$\lambda_0$ ו-$\omega_0$ אשר חייבים להיות מזווגים בגלל האילוץ, הרי שאת שאר הערכים העצמיים ניתן לסדר כרצוננו.

בנוסף מנוצלת העובדה שהמטריצות $Q_N$ ו-$\tilde{D}$ הן סימטריות והוקטור $\underline{1}$ הוא וקטור עצמי שלהן על מנת להראות ש-$\lambda_0$ הוא הערך העצמי הגדול ביותר (לפי Perron-Frobenius [La85], סעיף 15.3) של $Q_N$, ו-$\omega_0$ הוא הערך העצמי הגדול ביותר של $\tilde{D}$.

נסדר את הערכים העצמיים בסדר יורד ונסמן אותם ע"י $\lambda_i$, $i = 1,2,\dots,N-1$ וע"י $\omega_i$, $i = 1,2,\dots,N-1$. הערך המקסימלי יתקבל כאשר נבצע את הכפל ב-(6.9) לפי סדר זה. הערך המינימלי יתקבל כאשר נהפוך את הסדר באחת הקבוצות. כלומר:

$$\lambda_0\omega_0 + \sum_{i=1}^{N-1} \lambda_i \omega_{N-i} \le trace\{\Lambda U^T \tilde{D} U\} \le \lambda_0\omega_0 + \sum_{i=1}^{N-1} \lambda_i \omega_i \tag{6.10}$$

כאשר $\lambda_1 \ge \lambda_2 \ge \dots \ge \lambda_{N-1}$ וכן $\omega_1 \ge \omega_2 \ge \dots \ge \omega_{N-1}$. $\lambda_0 = 1$ הוא הערך העצמי של המטריצה $Q_N$ המתאים לוקטור העצמי $\underline{1}$ ו-$\omega_0$ הוא הערך העצמי של $\tilde{D}$ המתאים לאותו וקטור עצמי.

נציב את (5.11) ב-(5.8) ונקבל את החסמים על העיוות כתוצאה משגיאות ערוץ:

$$\frac{1}{2}\left(\lambda_0\omega_0 + \sum_{i=1}^{N-1} \lambda_i \omega_{N-i}\right) - S \le D_C \le \frac{1}{2}\left(\lambda_0\omega_0 + \sum_{i=1}^{N-1} \lambda_i \omega_i\right) - S \tag{6.11}$$

לסיכום, מציאת החסמים נעשית לפי ה-"מתכון" הבא (צעדים 1-3 הם עפ"י האלגוריתם שתואר בסעיף 6.2.2):

**<u>אלגוריתם למציאת החסמים על עיוות הערוץ מעל כל השיבוצים האפשריים</u>**

<u>נתונים</u>: מטריצת הערוץ $Q$ (2.11).

מטריצת המרחקים $D$ (2.14)

מטריצת ההסתברויות $P$ (2.14)

<u>אתחול</u>: מטריצת המרחקים המשוקללת $\tilde{D} \leftarrow \hat{D} = DP + P^T D^T$

סכום הסקלרים המתווספים $S \leftarrow 0$.

<u>צעד 1</u>: חשב את הסכום $S_i$ של כל השורות במטריצה $\tilde{D}$. $i = 0,1,\ldots,N-1$ , $S_i = \sum_{j=0}^{N-1} \tilde{D}_{ij}$.

<u>צעד 2</u>: חפש את הסכום המקסימלי וסמן את השורה עם סכום זה ב-$k$. $S_k = \max_{i=0,1,\ldots,N-1} S_i$.

<u>צעד 3</u>: עבור כל שורה $i \neq k$:

הוסף מטריצה במבנה צלב: $\tilde{D} \leftarrow \tilde{D} + \frac{1}{N}(S_k - S_i)(C_i + C_i^T)$.

עדכן את סכום הסקלרים: $S \leftarrow S + \frac{1}{N}(S_k - S_i)$.

<u>צעד 4</u>: חשב את הערכים העצמיים של מטריצת הערוץ $Q$, וסדר אותם בסדר יורד $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{N-1}$. עבור הערוץ הבינארי הסימטרי היעזר **בנספח ב'** (B.23).

<u>הערה</u>: $\lambda_0 = 1$ הוא הערך העצמי של $Q$ המתאים לוקטור $\underline{1}$.

<u>צעד 5</u>: חשב את הערכים העצמיים של מטריצת המרחקים המשוקללת $\tilde{D}$

וסדר אותם בסדר יורד $\omega_0 \geq \omega_1 \geq \cdots \geq \omega_{N-1}$.

<u>הערה</u>: $\omega_0$ הוא הערך העצמי של $\tilde{D}$ המתאים לוקטור $\underline{1}$.

<u>צעד 6</u>: חשב את החסמים מתוך:

$$\frac{1}{2}\left(\lambda_0\omega_0 + \sum_{i=1}^{N-1}\lambda_i\omega_{N-i}\right) - S \leq D_C \leq \frac{1}{2}\left(\lambda_0\omega_0 + \sum_{i=1}^{N-1}\lambda_i\omega_i\right) - S$$

בסעיף הבא ניתנות דוגמאות ספציפיות למציאת החסמים.

## 6.3 דוגמאות מספריות

בסעיף זה נחשב את החסמים עבור מקרים ספציפיים. בגלל מספרם הרב ($N!$) של השיבוצים האפשריים, הגרלנו עבור כל מקרה 10,000 שיבוצים אקראיים וחישבנו את עיוות הערוץ בכל אחד מהם. נערוך השוואה בין החסמים לבין השיבוץ הטוב ביותר והשיבוץ הגרוע ביותר שנמצאו בסימולציה. ממוצע הסימולציה יצוין אף הוא.

עבור שני מקרים ערכנו השוואה בין החסמים לבין השיבוץ המתקבל מאלגוריתם החלפת זוגות המתואר ב-[Ze87]. בנוסף ערכנו במקרים אלו השוואה עם השיבוץ המתקבל באלגוריתם המתואר בפרק הבא.

### 6.3.1 קוונטייזר סקלרי אחיד ומקור אחיד תחת ערוץ בינארי סימטרי

בסוף נספח ג׳ תחת הערה מספר 3, מופיע חישוב מפורט של החסמים עבור קוונטייזר סקלרי אחיד, מקור אחיד וערוץ בינארי סימטרי. החסמים עבור מקרה זה חושבו גם בפרק 5. סכום הסקלרים אשר הוספו למטריצה $\hat{D}$ לקבלת המטריצה $\tilde{D}$ הוא (C.18): $S = \dfrac{4(N-1)(N-2)}{3N^2}$.

הערכים העצמיים של המטריצה $\tilde{D}$ הם (C.21):

$$\begin{aligned}
\omega_0 &= \frac{4(N-1)^2}{N^2} \quad \Leftrightarrow \quad \underline{1} \\
\omega_1 &= \omega_2 = \cdots = \omega_{N-2} = 0 \\
\omega_{N-1} &= -\frac{4(N-1)(N+1)}{3N^2}
\end{aligned} \tag{6.12}$$

עבור הערוץ הבינארי הסימטרי הערכים העצמיים של המטריצה $Q_N$ כפי שחושבו בנספח ב׳ הם:

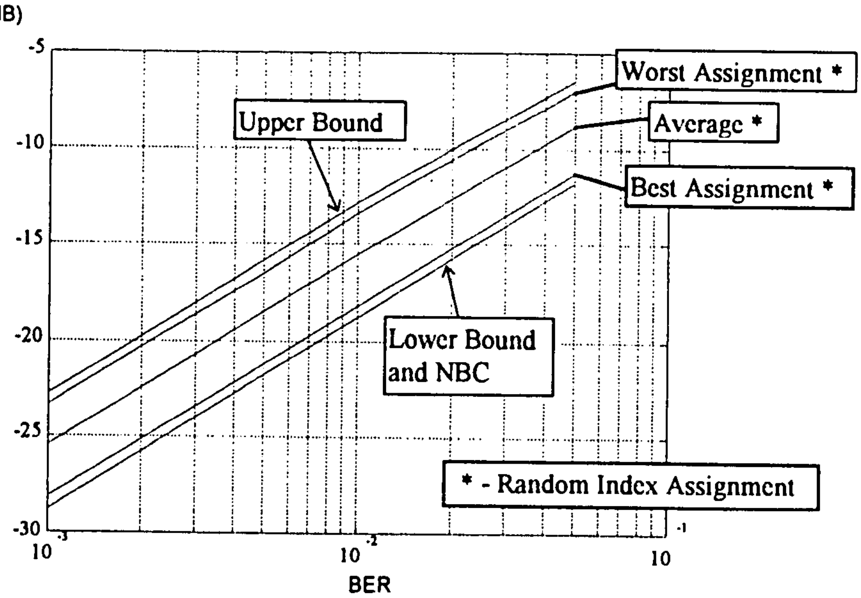
$$\begin{aligned}
\lambda_0 &= 1 \quad \Leftrightarrow \quad \underline{1} \\
\lambda_1 &= 1-2q \\
&\vdots \\
\lambda_{N-1} &= (1-2q)^L
\end{aligned} \tag{6.13}$$

הצבת הגדלים ב-(6.11) תיתן:

$$\frac{2(N+1)(N-1)}{3N^2}\cdot 2q \le D_c \le \frac{2(N+1)(N-1)}{3N^2}\left[1-(1-2q)^L\right] \tag{6.14}$$

כפי שנתקבל ב-(5.13).

בציור 6.1 מופיעים החסמים עבור קוונטייזר בן 4 סיביות, בהשוואה לתוצאות הסימולציה. השיבוץ האופטימלי, שהוא הקוד הבינארי הסימטרי (NBC), מתלכד במקרה זה עם החסם התחתון.




ציור 6.1 - חסמים, עליון ותחתון, מעל כל שיבוצי הקוד האפשריים, על עיוות-הערוץ עבור קוונטייזר סקלרי אחיד בן 4 סיביות ומקור אחיד תחת ערוץ בינארי סימטרי. להשוואה מוצגות תוצאות סימולציה בה נבחנו 10,000 שיבוצים אקראיים. בציור סומנו השיבוץ הטוב ביותר שנמצא, השיבוץ הגרוע ביותר וממוצע הסימולציה. במקרה זה ביצועי הקוד הבינארי הטבעי (NBC) מזדהים עם החסם התחתון.

Fig 6.1 - Upper and lower bounds over all possible index assignments on the Channel-Distortion of a 4-bit Uniform Scalar Quantizer and a uniform source under the Binary Symmetric Channel.
Bounds are compared with the best and worst assignments of 10,000 randomly picked assignments. The lower bound coincides with the performance of the optimal assignment (Natural Binary Code)

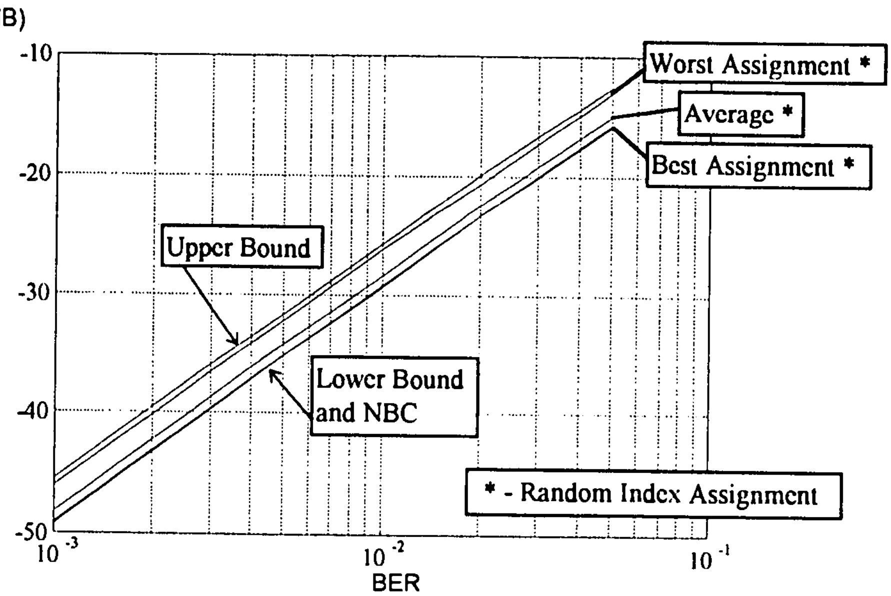
ניתן להבחין כי החסמים במקרה זה רחוקים כ-0.5dB בלבד משיבוצים שנתקבלו בסימולציה. השיבוץ האופטימלי למקרה זה (NBC) לא התקבל בתהליך הסימולציה.

לפי (5.14), המרחק בין החסמים עבור $L$ סיביות הוא $10 \cdot log_{10} L$ (dB), ועבור 4 סיביות כ-6dB. גם במקרה פשוט זה, ניתן לראות את החשיבות של שימוש מושכל בשיבוץ אינדקסים. הרווח עקב שיבוץ אינדקסים טוב צפוי לגדול עם העלייה במספר הסיביות.

## 6.3.2 קוונטייזר סקלרי אחיד ומקור אחיד תחת ערוץ בינארי סימטרי עם קוד Hamming (7,4) לתיקון שגיאות

ניתן להשתמש בטכניקת החסימה שהוצגה בפרק זה גם עבור ערוצים חסרי זיכרון השונים מן הערוץ הבינארי הסימטרי. נסתכל כעת על ערוץ הכולל גם קוד בלוק לתיקון שגיאות. עבור קוונטייזר בעל 16 רמות ייצוג, ניתן לשבץ לכל רמה אינדקס ערוץ בן 7 סיביות תוך שימוש בטבלת Hamming Code (7,4). במקרה זה המרחק בין מלות קוד הוא לפחות 3 סיביות, ולכן הקוד מסוגל לתקן שגיאה בסיבית אחת.

עבור ערוץ זה, נסתכל על המקרה של קוונטייזר סקלרי אחיד ומקור אחיד מסעיף 6.3.1. עדיין עומדת לרשותנו המטריצה $\tilde{D}$ והערכים העצמיים שלה. בציור 6.2 מופיעים החסמים עבור קוונטייזר בן 4 סיביות, בהשוואה לתוצאות הסימולציה.




ציור 6.2 - חסמים, עליון ותחתון, מעל כל שיבוצי הקוד האפשריים, על עיוות-הערוץ עבור קוונטייזר סקלרי אחיד בן 4 סיביות ומקור אחיד תחת ערוץ בינארי סימטרי וקוד Hamming (7,4). להשוואה מוצגות תוצאות סימולציה בה נבחנו 10,000 שיבוצים אקראיים. בציור סומנו השיבוץ הטוב ביותר שנמצא, השיבוץ הגרוע ביותר וממוצע הסימולציה. הקוד הבינארי הטבעי (NBC) מתלכד עם החסם התחתון.

Fig 6.2 - Upper and lower bounds over all possible index assignments on the Channel-Distortion of a 4-bit Uniform Scalar Quantizer and a uniform source under the Binary Symmetric Channel with (7,4) Hamming code.
Bounds are compared with the best and worst assignments of 10,000 randomly picked assignments. The Natural Binary Code coincide with the lower bound.

מהסתכלות בתוצאות הנומריות עולה כי במקרה זה החסם העליון רחוק כ-0.6dB מן השיבוץ הגרוע ביותר שנתקבל בסימולציה. החסם התחתון רחוק רק ב-0.1dB מן השיבוץ הטוב ביותר שנתקבל. המרחק בין החסמים עומד על 4.5dB בערך. ניתן לראות כי השימוש בקוד לתיקון שגיאות הוריד את ההפרש בביצועים בין שיבוץ "טוב" לבין שיבוץ "גרוע".

מעניין לציין כי בדומה לערוץ הבינארי הסימטרי, מטריצת הערוץ $Q$ ניתנת ללכסון ע"י מטריצת Walsh-Hadamard. הוכחה מפורטת של עובדה זו מופיעה **בנספח ד'**. וקטור רמות הייצוג $\underline{\rho}$ (5.1) הוא וקטור עצמי המתאים לערך העצמי הגדול ביותר, פרט ל-1. מסיבה זו הקוד הבינארי הטבעי (NBC) מתלכד עם החסם התחתון.

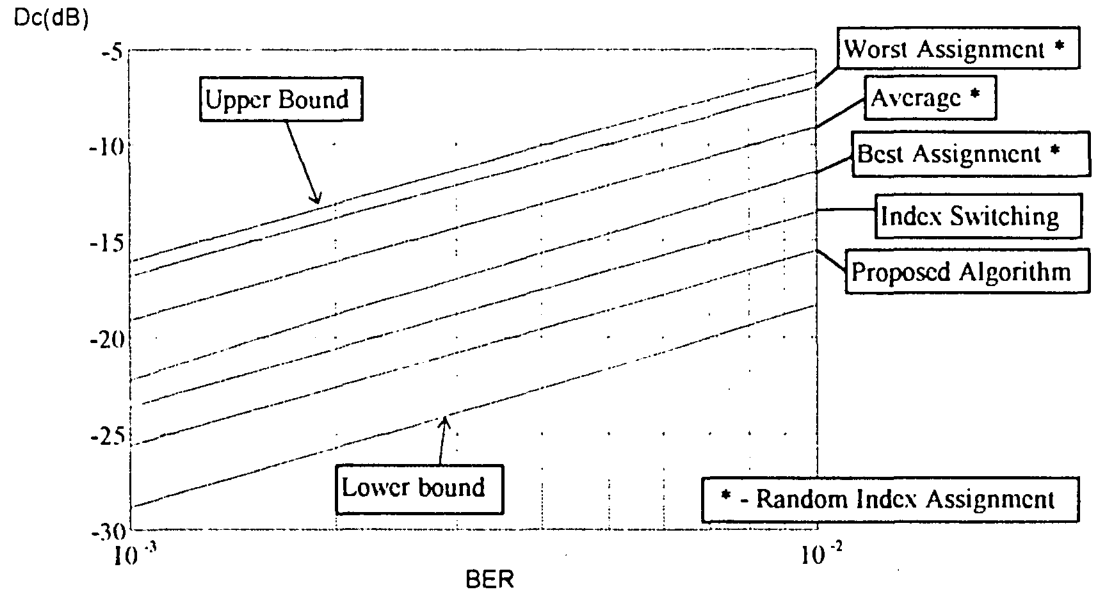
המסקנה היא <u>שגם עבור ערוץ זה הקוד הבינארי הסימטרי הוא השיבוץ האופטימלי</u>.

### <u>6.3.3 קוונטייזר אחיד מותאם למקור גאוסי תחת ערוץ בינארי סימטרי</u>

נתייחס לקוונטייזר אחיד בן 4 סיביות שצעד הקוונטיזציה שלו מותאם לפילוג גאוסי [Ja84 סעיף 4.3.2]. מטריצת הערוץ זהה לזו שבסעיף 6.3.1. למציאת החסם חישבנו את מטריצת המרחקים המשוקללת ואת הערכים העצמיים שלה:

$$\begin{aligned} \omega_0 &\approx 9.427 \quad ; \quad \omega_3 \approx \ldots \approx \omega_{13} \approx 0 \\ \omega_1 &\approx 1.248 \quad ; \quad \omega_{14} \approx -3.281 \\ \omega_2 &\approx 0.372 \quad ; \quad \omega_{15} \approx -4.131 \end{aligned} \tag{6.15}$$

בציור 6.3 מופיעים החסמים עבור קוונטייזר בן 4 סיביות, בהשוואה לתוצאות הסימולציה. בנוסף מופיעים בציור הביצועים עבור שיבוץ המתקבל מאלגוריתם החלפת זוגות המתואר ב-[Ze87]. בנוסף ערכנו השוואה עם השיבוץ המתקבל באלגוריתם לשיבוץ אינדקסים המתואר בפרק 7.



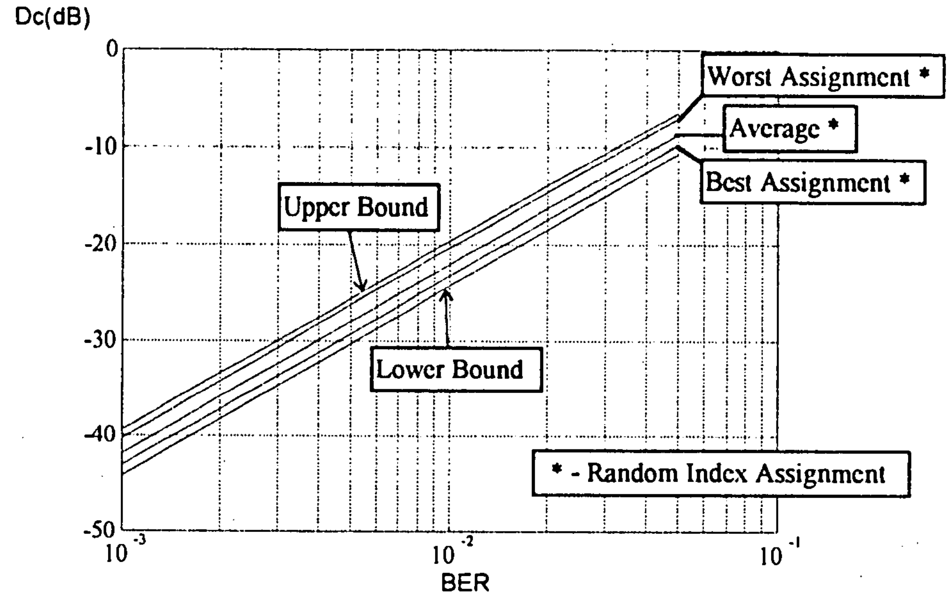

ציור 6.3 - חסמים, עליון ותחתון, מעל כל שיבוצי הקוד האפשריים, על עיוות-הערוץ עבור
קוונטייזר סקלרי אחיד בן 4 סיביות מותאם לפילוג גאוסי תחת ערוץ בינארי סימטרי.
להשוואה מוצגות תוצאות סימולציה בה נבחנו 10,000 שיבוצים אקראיים.
בציור סומנו השיבוץ הטוב ביותר שנמצא, השיבוץ הגרוע ביותר וממוצע הסימולציה.
בנוסף הובאו בציור הביצועים של השיבוצים המתקבלים מאלגוריתם החלפת אינדקסים
המתואר ב-[Ze87] ואלגוריתם המוצע בפרק 7.

F 6.3 - Upper and lower bounds over all possible index assignments on the Channel-Distortion of a 4-bit PDF optimized Uniform Scalar Quantizer and a Gaussian source under the Binary Symmetric Channel. Bounds are compared with the best and worst assignments of 10,000 randomly picked assignments. Bounds are also compared with the performance of Index Assignments obtained from Index swithing algorithm [Ze87] and the algorithm proposed in Chapter 7.

מהסתכלות בתוצאות הנומריות עולה כי במקרה זה החסם העליון רחוק כ-0.6dB מן השיבוץ הגרוע ביותר שנתקבל בסימולציה. החסם התחתון רחוק ב-6dB מן השיבוץ הטוב ביותר שנתקבל בסימולציה. המרחק בין החסמים עומד על 12dB בערך. ניתן לראות כי חיפוש אקראי של שיבוצים ($10^{4}$ שיבוצים מתוך $2\cdot10^{13}$ שיבוצים אפשריים) לא נתן מידע מספיק על הביצועים שניתן להשיג מן המערכת עבור שיבוצים שונים. השיבוץ שנתקבל באלגוריתם חילופי הזוגות הביא לשיבוץ טוב מזה שנמצא בחיפוש האקראי. השיבוץ שנתקבל מן האלגוריתם המתואר בפרק 7. נתן שיבוץ טוב עוד יותר.

### 6.3.4 קוונטייזר אחיד מותאם למקור גאוסי תחת ערוץ בינארי סימטרי עם קוד Hamming (7,4) לתיקון שגיאות

נתבונן בקוונטייזר מסעיף 6.3.3. עומדת לרשותנו מטריצת המרחקים המשוקללת מן הסעיף הקודם. בציור 6.4 מופיעים החסמים עבור קוונטייזר בן 4 סיביות, בהשוואה לתוצאות הסימולציה.



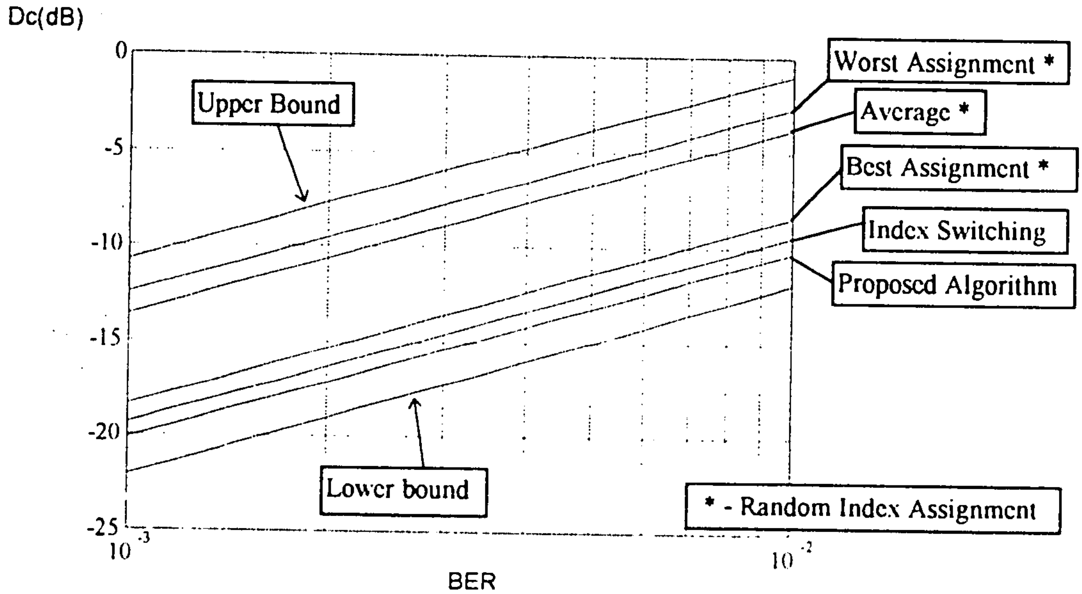

ציור 6.4 - חסמים, עליון ותחתון, מעל כל שיבוצי הקוד האפשריים, על עיוות-הערוץ עבור קוונטייזר סקלרי אחיד בן 4 סיביות מותאם לפילוג גאוסי תחת ערוץ בינארי סימטרי וקוד Hamming (7,4).
להשוואה מוצגות תוצאות סימולציה בה נבחנו 10,000 שיבוצים אקראיים.
בציור סומנו השיבוץ הטוב ביותר שנמצא, השיבוץ הגרוע ביותר וממוצע הסימולציה.

F 6.4 - Upper and lower bounds over all possible index assignments on the Channel-Distortion of a 4-bit PDF optimized Uniform Scalar Quantizer and a Gaussian source under the Binary Symmetric Channel with (7,4) Hamming code. Bounds are compared with the best and worst assignments of 10,000 randomly picked assignments.

מהסתכלות בתוצאות הנומריות עולה כי במקרה זה החסם העליון רחוק כ-0.9dB מן השיבוץ הגרוע ביותר שנתקבל בסימולציה. החסם התחתון רחוק ב-0.8dB מן השיבוץ הטוב ביותר שנתקבל. המרחק בין החסמים עומד על 3dB בערך. בדומה למקור האחיד גם כאן תוספת הקוד לתיקון שגיאות הורידה את ההפרש בין שיבוץ "טוב" לשיבוץ "גרוע".

### 6.3.5 קוונטייזר וקטורי מותאם למקור גאוסי תחת ערוץ בינארי סימטרי

לצורך סעיף זה השתמשנו בקוונטייזר וקטורי בן 8 סיביות, 2 סיביות לדגימה שתוכנן עבור מקור גאוס-מרקוב עם מקדם מתאם 0.5 (ראה גם ציור 3.3). בציור 6.5 מופיעים החסמים עבור הקוונטייזר, בהשוואה לתוצאות הסימולציה. הפעם נעשתה השוואה מול 100,000 שיבוצים אקראיים. בנוסף מופיעים בציור הביצועים עבור שיבוץ המתקבל מאלגוריתם החלפת זוגות המתואר ב-[Ze87]. בנוסף ערכנו השוואה עם השיבוץ המתקבל באלגוריתם לשיבוץ אינדקסים המתואר בפרק 7.




ציור 6.5 - חסמים עליון ותחתון מעל כל שיבוצי הקוד האפשריים על עיוות-הערוץ עבור קוונטייזר וקטורי בן 8 סיביות (2 סיביות לדגימה) מותאם לפילוג גאוסי (מקדם מתאם 0.5) תחת ערוץ בינארי סימטרי.
להשוואה מוצגות תוצאות סימולציה בה נבחנו 100,000 שיבוצים אקראיים.
בציור סומנו השיבוץ הטוב ביותר שנמצא, השיבוץ הגרוע ביותר וממוצע הסימולציה. בנוסף הובאו בציור הביצועים של השיבוצים המתקבלים מאלגוריתם החלפת אינדקסים המתואר ב-[Ze87] ואלגוריתם המוצע בפרק 7.

Fig 6.5 - Upper and lower bounds over all possible index assignments on the Channel-Distortion of a 8-bit Vector Quantizer (2 bit/sample) and a Gaussian source (correlation 0.5 ) under the Binary Symmetric Channel.
Bounds are compared with the best and worst assignments of 100,000 randomly picked assignments. Bounds are also compared with the performance of Index Assignments obtained from Index swithing algorithm [Ze87] and the algorithm proposed in Chapter 7.

ניתן לראות כי החסם העליון רחוק כ-2dB מן השיבוץ הגרוע ביותר שנמצא. החסם התחתון רחוק כ-3dB מהשיבוץ הטוב ביותר שנמצא בחיפוש. המרחק בין החסמים הוא כ-11dB. ההפרש בין השיבוץ הגרוע ביותר לשיבוץ הטוב ביותר הוא כ-6dB. תוצאה נומרית דומה מופיע ב-[Ze90].
גם בדוגמה זו ניתן לראות כי חיפוש אקראי של שיבוצים איננו מספק מבחינת מציאת שיבוץ אינדקסים יעיל. השיבוץ שנתקבל באלגוריתם חילופי הזוגות הביא לשיבוץ טוב מזה שנמצא בחיפוש האקראי. השיבוץ שנתקבל מן האלגוריתם המתואר בפרק 7. נתן שיבוץ טוב עוד יותר.

## 6.4 סיכום

בפרק זה הוצגו חסמים על עיוות הערוץ מעל כל השיבוצים האפשריים של מילות קוד לוקטורים מייצגים. החסמים יכולים להיות מיושמים על קוונטייזרים סקלריים ווקטוריים כלשהם. אין הגבלה גם על מידת המרחק ועל הסתברויות המקורות. הדרישה על הערוץ היא היותו חסר-זיכרון וסימטרי. כמו-כן הוצגו חסמים מספריים עבור מקרים ספציפיים.
החסמים הינם הרחבה של התוצאות מפרק 5, שם נמצאו חסמים עבור המקרה המיוחד של קוונטייזר סקלרי אחיד, מקור אחיד וערוץ בינארי סימטרי ונמצא כי השיבוץ הבינארי הטבעי (NBC) הוא השיבוץ האופטימלי. בפרק זה הוכחנו כי הקוד הבינארי הטבעי הוא השיבוץ האופטימלי עבור מקרה נוסף: קוונטייזר סקלרי אחיד, מקור אחיד וערוץ בינארי סימטרי עם קוד Hamming (7,4) לתיקון שגיאות.
השוואה נוספת נעשתה עם 10,000 קודים שנבחרו באקראי (כאשר עבור קוונטייזר וקטורי עם 8 סיביות נעשתה השוואה עם 100,000 קודים).
כאשר השווינו ביצועים של קוונטייזר הפועל תחת הערוץ הבינארי הסימטרי, ראינו בדוגמאות המספריות כי תוספת הקוד לתיקון שגיאות הורידה את ההפרש בין הביצועים של שיבוצים טובים לשיבוצים גרועים.
השווינו את החסמים לביצועים עם שיבוץ המתקבל מאלגוריתם של חילופי זוגות וכן עם השיבוץ לפי האלגוריתם המוצע בפרק 7. במקרים שנבדקו נתן השיבוץ שהתקבל מן האלגוריתם הראשון ביצועים טובים מחיפוש אקראי בעוד שהאלגוריתם המתואר בפרק 7 נתן ביצועים טובים עוד יותר.

# פרק 7 - אלגוריתם תת-אופטימלי למציאת שיבוץ אינדקסים עבור ערוץ סימטרי חסר זיכרון

## 7.1 מבוא

בפרקים הקודמים עמדנו על מספרם העצום של השיבוצים האפשריים של מילות קוד לוקטורי-ייצוג בקוונטייזרים וקטוריים. ראינו בפרק 5 כי במקרה המיוחד של קוונטייזר סקלרי אחיד, מקור אחיד וערוץ בינארי סימטרי, ניתן להצביע על השיבוץ האופטימלי, שהוא הקוד הבינארי הטבעי. בפרק 6 תוארו חסמים על עיוות הערוץ מעל כל $N!$ השיבוצים האפשריים של מילות קוד בקוונטייזרים וקטוריים כלשהם. החסמים ניתנים ליישום עבור ערוצים <u>חסרי-זיכרון סימטריים</u>. בפרק זה נתאר אלגוריתם קונסטרוקטיבי תת-אופטימלי למציאת שיבוץ אינדקסים עבור קוונטייזר וקטוריים הפועלים תחת ערוץ סימטרי חסר זיכרון.
פיתוח האלגוריתם מופיע בשלמותו **בנספח ה'**, בפרק זה נציג את תקציר הפיתוח בלבד. האלגוריתם מבוסס על תוצאות שנתנו בפרק 6. הסיבוכיות של האלגוריתם היא $O(N^3)$. בדוגמאות מספריות שהוצגו בפרק 6, השווינו את ביצועי האלגוריתם המוצע עם ביצועי אלגוריתם חילופי-אינדקסים, בעל סיבוכיות $O(N^4)$. ראינו כי ביצועי האלגוריתם המוצע טובים מביצועי האלגוריתם של חילופי-אינדקסים.

## 7.2 תקציר פיתוח האלגוריתם

לפי (2.15), ביטוי כללי לעיוות הערוץ ניתן ע"י:

$$D_c = trace\left\{P\pi Q_N \pi^T D\right\} \tag{7.1}$$

אנו מניחים כי מטריצת הערוץ $Q_N$ הינה סימטרית. מכיוון שהיא מייצגת הסתברויות, סכום כל שורה שווה לאחד ולכן $Q_N \cdot \underline{1} = \underline{1}$. אין הנחות נוספות על מטריצת ההסתברויות $P$ (אלכסונית) ועל מטריצת המרחקים בין הווקטורים המיצגים $D$. השיבוץ מתואר ע"י מטריצת הפרמוטציה $\pi$.

בפרק 6 ראינו כי בעיית מציאת המינימום לעיוות הערוץ שקולה לפתרון הבעיה:

$$\min_{\pi} \quad trace\left\{Q_N \pi^T \tilde{D} \pi\right\} \tag{7.2}$$

כאשר מטריצת המרחקים המשוקללת $\tilde{D}$ נתקבלה מן המטריצות $P$ ו-$D$ ע"י אלגוריתם המתואר בסעיף 6.2.2 ובנספח ג'. המטריצה $\tilde{D}$ היא סימטרית ומקיימת את התכונה כי הוקטור $\underline{1}$ הוא וקטור עצמי שלה עם ערך עצמי מתאים $\omega_0$.

נבצע לכסון יוניטרי למטריצת הערוץ הסימטרית $Q_N$ ולמטריצת המרחקים המשוקללת $\tilde{D}$.

$$
\begin{aligned}
Q_N &= V \cdot \Lambda \cdot V^T \qquad & V \cdot V^T = I \\
\tilde{D} &= Z \cdot \Omega \cdot Z^T \qquad & Z \cdot Z^T = I
\end{aligned}
\tag{7.3}
$$

נסדר את הערכים העצמיים של המטריצות בסדר יורד. כאשר הערך העצמי הראשון של $Q_N$ הוא $\lambda_0 = 1$ והוקטור העצמי המתאים הוא $\frac{1}{\sqrt{N}} \cdot \underline{1}$. באופן דומה, הערך העצמי הראשון של $\tilde{D}$ הוא $\omega_0$ והוקטור העצמי המתאים הוא $\frac{1}{\sqrt{N}} \cdot \underline{1}$. ראינו בסעיף 6.2.4 כי עבור מטריצות סימטריות אי-שליליות בעלות וקטור עצמי השווה ל- $\underline{1}$, הערך העצמי המתאים הוא הערך העצמי הגדול ביותר בערך מוחלט (לפי Perron-Frobenius [La85], סעיף 15.3).

נציב את (7.3) ב-(7.2):

$$
\begin{aligned}
&\min_{\pi} \; trace\left\{ V \cdot \Lambda \cdot V^T \pi^T Z \cdot \Omega \cdot Z^T \pi \right\} = \\
&\min_{\pi} \; trace\left\{ \Lambda \cdot V^T \pi^T Z \cdot \Omega \cdot Z^T \pi \cdot V \right\} = \\
&\qquad\qquad \Leftarrow trace\left\{ AB \right\} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} A_{ij}^T B_{ij} \\
&\min_{\pi} \; \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \lambda_j \omega_i \left( Z^T \pi V \right)_{ij}^2 = \\
&\min_{\pi} \; \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \lambda_j \omega_i \Psi_{ij}^2
\end{aligned}
\tag{7.4}
$$

המטריצה $\Psi = Z^T \pi V$ היא יוניטרית מכיוון ש:

$$
\Psi \Psi^T = \left( Z^T \pi V \right) \left( Z^T \pi V \right)^T = Z^T \pi V V^T \pi^T Z = Z^T \pi \pi^T Z = Z^T Z = I \tag{7.5}
$$

מכיוון שהעמודה הראשונה של $V$ והעמודה הראשונה של $Z$ הן $\underline{v}_0 = \underline{z}_0 = \frac{1}{\sqrt{N}} \underline{1}$, ושאר העמודות אורתוגונליות לוקטור $\underline{1}$, הרי שלמטריצה $\Psi$ יש את המבנה הבא:

$$\Psi = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & ? & \\ 0 & & & \end{bmatrix} \tag{7.6}$$

כאשר סימן השאלה מתאר איברים לא ידועים.

בנספח ג' (C.12) ראינו כי אם המטריצה היוניטרית $\pi V$ מוחלפת במטריצה יוניטרית כללית, $U$, אזי הערך המינימלי של פונקצית המטרה ב-(7.4) מתקבל עבור:

$$\Psi^{*} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & & & 1 \\ \vdots & & & 1 & \\ 0 & & \cdot^{\cdot^{\cdot}} & & \\ 0 & 1 & & & 0 \end{bmatrix} \tag{7.7}$$

המטריצה $\Psi^{*}$ איננה תוצאה של מטריצת פרמוטציה חוקית בביטוי $\Psi = Z^{T}\pi V$, ולכן נבצע קירוב של $\pi$ ע"י מטריצה יוניטרית כללית. אם נחליף את מטריצת הפרמוטציה $\pi$ במטריצה יוניטרית כללית $W$, הערך המינימלי של פונקצית המטרה יתקבל עבור $W = Z\Psi^{*}V^{T}$. יש לשים לב כי מכיוון ש-$W$ יוניטרית, כל איברי המטריצה קטנים בערכם המוחלט מ-1, $\left|W_{ij}\right| \le 1 \,,\quad i,j = 0,1,\ldots,N-1$.

נחפש כעת מטריצת פרמוטציה "קרובה" למטריצה $W$, ע"י פתרון בעיית שיבוץ רגילה הידועה מתורת חקר ביצועים.

$$\begin{aligned} \min_{\pi} \quad & \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(1-w_{ij}^{2}\right)\pi_{ij} \\ s.t. \quad & \sum_{j=0}^{N-1}\pi_{ij} = 1 \quad i = 0,1,\ldots,N-1 \\ & \sum_{i=0}^{N-1}\pi_{ij} = 1 \quad j = 0,1,\ldots,N-1 \\ & \pi_{ij} \ge 0 \end{aligned} \tag{7.8}$$

פונקצית המשקל $\left(1-w_{ij}^{2}\right)$ הנה אחת מפונקציות משקל אפשריות, אשר נבחרה כי היא נמצאה יעילה בפתרון המקרים הספציפיים אותם בדקנו.

פתרון בעיית השיבוץ (7.8), יכול להיעשות בשיטה ההונגרית [Ta87 סעיף 6.3].
התוצאה של האלגוריתם היא מטריצת פרמוטציה $\pi$, בה ניתן להשתמש בבעיה המקורית (7.1).

האלגוריתם מורכב מן השלבים הבאים:

1. מציאת ערכים עצמיים של שתי מטריצות הרמיטיות [Go89 סעיף 8.2]: $O(N^3)$.
2. כפל מטריצות: $O(N^2)$.
3. פתרון בעית שיבוץ סטנדרטית [Ta87 סעיף 6.3]: $O(N^3)$.

סה"כ הסיבוכיות של האלגוריתם היא $O(N^3)$.

## 7.3 שימוש באלגוריתם למקרים אסימפטוטיים

בסעיף 3.4 דנו בערוץ הבינארי הסימטרי ובערוצים הנובעים ממנו ע"י קוד בלוק לתיקון שגיאות, כדוגמת קוד Hamming. איברי מטריצת הערוץ $Q$ תלויים בהסתברות השגיאה בערוץ הבינארי הסימטרי - $q$. נסמן עובדה זו באופן מפורש ע"י $Q(q)$. מכיוון שאיברי האלכסון במטריצת הערוץ מתארים קליטה ללא שגיאה, הגדרנו ב-(3.4) מטריצה בעלת אלכסון המכיל אפסים:

$$\tilde{Q}_{ij}(q)=\begin{cases} Q_{ij}(q) & i\neq j \\ 0 & i=j \end{cases} \tag{7.9}$$

חיפשנו את גורם השגיאה הדומיננטי בשאר איברי מטריצת הערוץ. המטריצה $\tilde{Q}(q)$ מורכבת מסכומים של איברים מן הצורה $l=L^0,L^0+1,\dots,L$ ,$q^l(1-q)^{L-l}$, כאשר סימנו את החזקה הנמוכה ביותר ב-$L^0$. הסיבה לכך שאיפסנו את האלכסון של המטריצה המקורית $Q(q)$היא שהיו באלכסון איברים בחזקות נמוכות יותר מ-$L^0$. נבחן את הגבול:

$$\lim_{q\to 0}\frac{1}{q^{L^0}}\cdot\tilde{Q}(q)=\tilde{Q}_A \tag{7.10}$$

איברים עם חזקות של $q$ הגבוהות מ-$L^0$ יתאפסו בגבול. קצב השגיאות בערוץ איננו מופיע לכן במטריצה $\tilde{Q}_A$. שימוש במטריצה $\tilde{Q}_A$ בבעיה המקורית (7.1) יביא לשיבוץ המתאים לתחום של קצבי שגיאה בערוץ מאפס שגיאות (ואז השיבוץ איננו חשוב) ועד לקצב $q_0$ בו הקירוב האסימפטוטי איננו מתאים. ראינו בפרק 3 כי ההתנהגות האסימפטוטית של עיוות הערוץ קרובה מאד להתנהגות האמיתית מהסתברויות שגיאה בערוץ של $1-5\%$ ומטה.

## 7.4 סיכום

בפרק זה הוצג אלגוריתם תת-אופטימלי למציאת שיבוץ של אינדקסים לוקטורי ייצוג בקוונטייזרים וקטוריים הפועלים תחת ערוץ סימטרי חסר זיכרון. הראינו כי לאלגוריתם סיבוכיות $O(N^3)$ בהשוואה לסיבוכיות $O(N^4)$ של אלגוריתם חילופי-אינדקסים המתואר בספרות. הצגנו גם את הגישה לשיבוץ עבור המקרה האסימפטוטי של קצבי שגיאה נמוכים בערוץ. עבור קצבי שגיאה נמוכים ($q_0 \approx 1-5\%$ ומטה) מתקבל שיבוץ טוב לכל התחום מאפס שגיאות (ואז השיבוץ איננו חשוב) ועד לקצב $q_0$.

השיבוץ שנתקבל עבור מקרים ספציפיים תואר בפרק 6, שם הראינו כי חיפוש אקראי של מספר מוגבל של שיבוצים רחוק מן השיבוץ שהוצג כאן. בנוסף השווינו בשני מקרים את הביצועים המתקבלים מאלגוריתם של חילופי-אינדקסים לביצועי האלגוריתם המוצע. האלגוריתם המוצע נמצא עדיף בשני המקרים.

# פרק 8 - סיכום, מסקנות והצעות להמשך מחקר

## 8.1 סיכום

עבודת מחקר זו עוסקת בחקר הביצועים של מערכות קידוד המבוססות על קוונטיזציה וקטורית, כאשר המידע הספרתי מועבר דרך ערוצי תקשורת רועשים ובדרכים לשיפור הביצועים של מערכות אלו.

בהצגת הבעיה ראינו כי ניתן להתייחס אל העיוות הכולל כמורכב מעיוות קוונטיזציה הנגרם מייצוג האות האנלוגי (או אות דגום ברזולוצית אמפליטודה גבוהה) ע"י מידע ספרתי ומעיוות ערוץ הנגרם כתוצאה משגיאות במידע הספרתי שהועבר בערוץ. בעבודה עסקנו אך ורק בערוצים חסרי זיכרון.

בחלקה הראשון העבודה מפותחים חסמים על עיוות הערוץ <u>מעל כל מקורות האות האפשריים</u> במערכות נתונות. חסמים אלו מספקים למתכנן מידע כמותי על עיוות הערוץ. פיתוח החסמים מבוסס על תכונות הידועות מתורת התכנות הליניארי. הוצגו מספר דוגמאות מספריות להדגמת החסמים. ראינו כי עבור המקרה המיוחד של קוונטייזר סקלרי אחיד, שיבוץ אינדקסים בקוד בינארי טבעי וערוץ בינארי סימטרי התלכדו החסמים בהסתברויות שגיאה בערוץ, $q$, של עד 5%. בדוגמאות אחרות, המרווח בין החסם העליון לחסם התחתון הוא גדול יחסית וידיעתם איננה תורמת מספיק להערכת ביצועי המערכת. עבור מקרים אלו מוצגים חסמים הדוקים יותר עבור משפחה מצומצמת יותר של מקורות, המאופיינת ע"י מומנטים של הסתברות המקור ובפרט אילוץ על הספק המקור. כמו-כן מוצגים חסמים אסימפטוטיים עבור ערוץ בינארי סימטרי (בלי או עם קוד לתיקון שגיאות) כאשר הסתברות השגיאה בערוץ שואפת לאפס, $q \to 0$. ההתנהגות האסימפטוטית של עיוות הערוץ מאופיינת בשיפוע התלוי באופי הערוץ. למשל, עבור ערוץ בינארי סימטרי, הראינו כי ללא קוד לתיקון שגיאות עיוות הערוץ תלוי ביחס ישר להסתברות השגיאה בערוץ $q$, ואילו עבור קוד המתקן שגיאה בסיבית אחת, עיוות הערוץ תלוי ביחס ישר ל-$q^2$. הראינו גם שעבור קוונטייזר סקלרי אחיד בן $L$ סיביות, גודלו של עיוות הערוץ כמעט ואינו תלוי ב-$L$, כאשר משתמשים בשיבוץ הבינארי הטבעי. יתירה מזו, עבור $L \geq 4$, עיוות הערוץ עבור השיבוץ הבינארי הטבעי קטן יותר ממוצע העיוות מעל כל השיבוצים האפשריים, עבור כל מקור שהוא.

בנוסף למציאת חסמים על ביצועי מערכות קוונטיזציה וקטורית נתונות, העבודה מתרכזת בשני גורמים המשפיעים על עיוות הערוץ. הגורם הראשון הוא מבנה הקוונטייזר הוקטורי (והסקלרי כמקרה פרטי) מבחינת תאי חלוקה ווקטורי ייצוג. הגורם השני הוא שיבוץ האינדקסים המשודרים בערוץ לוקטורי הייצוג.

קוונטייזר וקטורי המתוכנן ללא התחשבות בשגיאות ערוץ מבוסס בדרך כלל על שני תנאים הכרחיים לאופטימליות: תנאי השכן הקרוב ותנאי הצנטרואיד. כאשר מתחשבים בשגיאות ערוץ, הוצגו בספרות תנאים הכרחיים אלטרנטיביים: תנאי השכן הקרוב המשוקלל ותנאי הצנטרואיד המשוקלל. תנאים אלו מביאים לתכן מסובך של הקוונטייזר ולמימוש יקר מבחינת דרישות זיכרון. בעבודה מוצגת גישה תת-אופטימלית המשתמשת בקוונטייזר שתוכנן ללא התחשבות

בשגיאות הערוץ, תוך ביצוע כיווץ ליניארי של תאי החלוקה ווקטורי הייצוג. הגישה שהוצגה היא פשוטה לתכנון וזולה למימוש, תוך תשלום לא גדול בביצועים בהשוואה לאלגוריתמים האופטימליים.

ראינו כי עיוות הערוץ מושפע משיבוץ מילות הקוד לוקטורי הייצוג של הקוונטייזר. מספרם הרב של השיבוצים האפשריים ($N!$ עבור קוונטייזר בעל $N$ וקטורי ייצוג) איננו מאפשר בדרך כלל חיפוש ממצה ע"י בדיקה של כולם. בעבודה פותחו חסמים על עיוות הערוץ <u>מעל כל השיבוצים האפשריים</u> של אינדקסים לוקטורי ייצוג עבור ערוץ חסר זיכרון. חסמים אלו מאפשרים למתכנן להעריך את השפעת השיבוץ על המערכת שבתכנון.

עבור המקרה המיוחד של קוונטייזר סקלרי אחיד ומקור אחיד הפועל בערוץ בינארי סימטרי ניתנת הוכחה כי הקוד הבינארי הטבעי הוא השיבוץ האופטימלי.

בנוסף, עבור ערוצים סימטריים חסרי זיכרון, קוונטייזר ומקור כלשהם, מפותח בעבודה אלגוריתם תת-אופטימלי למציאת שיבוץ יעיל של מילות קוד לוקטורי הייצוג בקוונטייזר. סיבוכיות האלגוריתם היא $O(N^3)$. בהשוואות מספריות נותן האלגוריתם שיבוץ בעל ביצועים עדיפים על השיבוץ המתקבל מאלגוריתם של חילופי אינדקסים המוצג בספרות, על אף שסיבוכיותו גבוהה יותר $O(N^4)$.

בעזרת סימולציות מחשב, נעשתה השוואה בין החסמים על עיוות הערוץ מעל כל השיבוצים האפשריים, חיפוש בעשרת אלפים (מאה אלף באחד המקרים) שיבוצים אקראיים ועבור שני מקרים גם עם השיבוץ התת-אופטימלי המוצע. מן ההשוואות עולה כי החיפוש האקראי איננו מביא לתוצאות מספקות בגלל העושר הרב של שיבוצים אפשריים (עבור קוונטייזר של 4 סיביות בלבד קיימים $2\cdot10^{13}$ שיבוצים אפשריים). עבור המקרה המיוחד של קוונטייזר סקלרי אחיד ושיבוץ בינארי טבעי, מתלכד השיבוץ המוצע עם החסם התחתון. במקרה טיפוסי שנבדק בסימולציה, נמוך עיוות הערוץ עבור השיבוץ האופטימלי ב-8dB בהשוואה לממוצע של עיוות הערוץ מעל כל השיבוצים האפשריים. עבור אותו מקרה ביצועי אלגוריתם חילופי האינדקסים גרועים ב-1dB מן האלגוריתם המוצע.

## <u>8.2 הצעות להמשך מחקר</u>

- בעבודה זו התמקדנו בשגיאות ערוץ עבור ערוצים חסרי זיכרון. הרחבתן של התוצאות ובעיקר נושאי השיבוץ לערוצים בעלי זיכרון הינה בעלת חשיבות למערכות מעשיות.

- הרחבה של הגישה התת-אופטימלית של שינוי ליניארי במבנה הקוונטייזר להקטנת השפעתן של שגיאות ערוץ ע"י שימוש בפונקצית כיווץ לא ליניארית. חסרונה העיקרי של שיטת הכיווץ הליניארי היא בהגדלת רעש ההעמסה (Overload Noise) עבור כיווץ משמעותי של הקוונטייזר. עיוות לא ליניארי אשר יבצע כיווץ חזק יחסית במרכז לעומת קצות התחום, יתכן ויביא לביצועים

טובים יותר. יתרונה של גישה זו הוא שגם עם פונקצית כיווץ מסובכת יותר, התכן עדיין מבוסס על קוונטייזר שתוכנן ללא התחשבות בשגיאות ערוץ ואין צורך לתכנן ולשמור ספריית קוונטייזרים.
גישה אפשרית נוספת היא, שאם איבר אחד בוקטורים המיועדים לקידוד מציג התנהגות שונה מן היתר, ייתכן וכדאי למצוא לו פונקצית כיווץ אינדבידואלית. במקרה זה נקבל בעייה בשני משתני כיווץ $\min_{\alpha_1,\alpha_2} D(\alpha_1, \alpha_2)$ ויש לחפש מינימום בפונקציה הדו-מימדית. ניתן, כמובן, לבצע הרחבה ליותר משני מימדים.

- הרחבה של בעיית שיבוץ האינדקסים גם למערכות קידוד מקור-ערוץ משולבות. לדוגמה, עבור קוונטייזר וקטורי בן $N = 2^L$ וקטורי ייצוג, ניתן להשתמש בערוץ עם אלפבית בגודל $M = 2^{L_1} > N$. הסיביות הפנויות יכולות לשמש לתיקון שגיאות. בעיית השיבוץ כעת היא בסיבוכיות גבוהה עוד יותר. יש להחליט לאילו וקטורי ייצוג ישובצו אילו אינדקסים (כאשר $M - N$ אינדקסים פנויים), ואילו וקטורים מייצגים יפוענחו במקלט לכל אינדקס ערוץ שנקלט. ראוי לציין כי במאמר שתואר בסקר הספרות [Fa87] ההתייחסות לשיבוץ שונה מן הגישה שלנו.

- בדיקת האופטימליות של הקוד הבינארי הטבעי עבור קודים אחרים לתיקון שגיאות (מלבד הקוד הספציפי שנבדק (Hamming (7,4)).

# References מקורות ספרות


[Be92] Ben-David G. and Malah D., "On the Performance of a Vector-Quantizater under Channel Errors", *Signal Processing VI Theories and applications, Proceeding of EUSIPCO-92*, Elsevier Amsterdam, pp. 1685-1688, Aug. 1992.

[Be93] Ben-David G. and Malah D.,"Properties of the Natural Binary Code Assignment for Uniform Scalar Quantizers under Channel Errors", *Proc. of the ECCTD-93*, Elsevier Amsterdam, pp. 773-778, Sep. 1993.

[Be93] Ben-David G. and Malah D.,"Simple Adaptation of Vector-Quantizers to Combat Channel Errors", *Proc. of the IEEE DSP-94 Workshop*, IEEE, Pisctaway NJ, pp. 41-44, Oct. 1994.

[Bu79] Burkard E., "Travelling Salesman and Assignment Problems: A Survey", *Annal of Discrete Mathematics 4*, North-Holland publishing - Amsterdam, pp. 193-215, 1979.

[Ch87] Chen J.H., Davidson G., Gersho A. and Zeger K.,"Speech Coding for the Mobile Satellite Experiment", *Proc. IEEE Int. Conf. on Communication*, Seattle, WA pp. 756-763, June 1987.

[Ch93] Chaing D.M. and Potter L., "Minimax Non-redundant Channel-Coding for Vector Quantization", *Proc. ICASSP-93*, Minneapolis, Minnesota, pp. 617-620, April 1993.

[Cl85] Clarke R.J., *Transform Coding of Images*, Academic-Press Lonodon, 1985.

[Cr69] Crimmins T.R., Horwitz H.M, Palermo C.J and Palermo R.V., "Minimization of Mean-Square Error for Data Transmitted Via Group Codes", *IEEE Trans. Inform. Theory*, vol. 15, no. 1, pp. 72-78, Jan. 1969.

[Cu94] Cuperman V., Liu F.H. and Ho P., "Robust Vector Quantization for Noisy Channels Using Soft Decision and Sequential Decoding", *European Tran. Telecomunication*, vol. 5, no. 5, pp.7/541-18/552, Oct. 1994.

[De87] De Marca J.R.B and Jayant N.S.,"An algorithm for Assignning Binary Indices to the Codevectors of Multi-Dimensional Quantizer", *Proc. IEEE Intl. Conf. Comm.*, Seattle, WA, pp. 1128-1132, June 1987.

[Fa87] Farvardin N. and Vaishampayan V., "Optimal Quantizer Design for Noisy Channels: An Approach to Combined Source-Channel Coding", *IEEE Trans. Inform. Theory*, vol. 33, no. 6, pp. 827-838, Nov. 1987.

[Fa90] Farvardin N., "A Study of Vector Quantization for Noisy Channels", *IEEE Trans. Inform. Theory*, vol. 36, no. 4, pp. 799-809, July 1990.

[Fa91] Farvardin N. and Vaishampayan V., "On the Performance and Complexity of Channel-Optimized Vector Quantizers", *IEEE Trans. Inform. Theory*, vol. 37, no. 1, pp. 155-159, Jan. 1991.

[Fl64] Fleischer P.E., "Sufficient Conditions for Achieving Minimum Distortion in a Quantizer", *1964 IEEE Intl. Conv. Rec.*, pp. 104-111.

[Ge92] Gersho A. and Gray R.M., *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Boston, 1992.

[Gi62] Gilmore P.C., "Optimal and Suboptimal Algorithms for the Quadratic Assignment Problem", *J. Soc. Indust. Appl. Math.*, vol. 10, no.2, pp. 305-313, June 1962.

[Go89] Golub G.H. and Van-Loan C.F, *Matrix Computations*, The John Hopkins University Press, Baltimore, 1989.

[Ha70] Hall K.M.,"An r-Dimensional Quadratic Placement Algorithm", *Management Science*, The Instiute of Management Sciences, vol. 17, pp. 219-229, Nov. 1970.

[Ja84] Jayant N.S. and Noll P. *Digital Coding of Waveforms*, Prentice-Hll, Englewood Cliffs, NJ, 1984.

[Ko90] Kohonen T., "The Self-Organizing Map", *Proc. IEEE*, vol. 78, no. 9, pp. 1464-1480, Sep. 1990.

[Kn92] Knagenhjelm P., "A recursive Design Method for Robust Vector Quantization", *Proc. Intl. Conf. Signal Processing Applications and Technology*, Boston, pp.948-954, Nov. 1992.

[Kn93] Knagenhjelm P., "How good is your Index Assignment?", *Proc. IEEE - ICASSP-93*, Minneapolis, Minnesota, pp. 423-426, April 1993.

[Kn95] Knagenhjelm P. and Erik Agrell, "The Hadamard Transform - A Tool for Index Assignment", Submitted to *IEEE Trans. Inform. Theory*.

[Ku69] Kurtenbach A.J. and Wintz P.A, "Quantizing for Noisy Channels", *IEEE Trans. Comm. Tech.*, vol. 17, no. 2, pp. 291-302, April 1969.

[La85] Lancaster P. and Tismenetsky M., *The Theory of Matrices*, Academic-Press, Orlando, 1985.

[Li80] Linde Y., Buzo A. and Gray R.M., "An algorithm for Vector Quantization Design", *IEEE Trans. Comm.*, COM-28, pp.84-95, Jan. 1980.

[Ll82] Lloyd S.P., "Least squares Quantization in PCM", *IEEE Trans. Inform. Theory*, IT-28, pp. 127-135, March 1982.

[Lu84] Luenberger D.G., *Linear and Nonlinear Programming*, Addison-Wesley, Reading, MA, 1984.

[Mc92] McLaughlin S.W., Ashley J.J and Neuhoff D.L, "The optimality of the Natural Binary Index Assignment", *Proc. of the Joint Dimacs/IEEE workshop on Coding and Quantization*, October 1992.

[Pa82] Papadimitriou C.H. and Steiglitz K., *Combinatorial Optimization: Algorithms and Complexity*, Prentice-Hll, Englewood Cliffs, NJ, 1982.

[Ry76] Rydbeck N. and Sundnerg C.E.W., "Analysis of Digital Errors in Nonlinear PCM Systems", *IEEE Trans. Comm.*, COM-24, pp. 59-65, Jan. 1976.

[Ta86] Taub H. and Schilling D.L.,*Principles of Communication Systems*, McGraw-Hill, New-York, 1986.

[Ta87] Taha H.A., *Operations Research - An Introduction*, Maxwell-Macmillan, New-York, 1987.

[To67] Totty R.E. and Clark G.C, "Reconstruction Error in Waveform Transmission", *IEEE Trans. Inform. Theory*, IT-13 pp. 336-338, April 1967.

[Vi79] Viterbi A.J. and Omura J.K, *Principles of Digital Communication and Coding*, McGraw-Hill, New-York, 1979.

[Ze87] Zeger K. and Gersho A., "Zero Redundency Channel Coding in Vector-Quantization", *Electronics Letters*, vol. 23, no. 12, pp. 654-656, June 1987.

[Ze88] Zeger K. and Gersho A., "Vector Quantization Design for Memoryless Noisy Channels", *Proc. Intl. Conf. Comm.*, Philadephia, PA, pp. 1593-1597, June 1988.

[Ze90] Zeger K. and Gersho A., "Pseudo-Gray Coding", *IEEE Trans. Comm.*, vol. 38, no. 12, pp. 2147-2158, Dec. 1990.

[Ze94] Zeger K. and Manzella V., "Asymptotic Bounds on Optimal Noisy Channel Quatization Via Random Coding", *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1926-1938, Nov. 1994.

# Appendix A נספח א'

## תכונות מטריצת המרחקים של הקוונטייזר הסקלרי האחיד

## Properties of the uniform scalar quantizer distance matrix

In this appendix we introduce some properties of the distance matrix associated with the uniform scalar quantizer. These properties will be used in **Apendices B and** C.

**Definition**: The uniform scalar quantizer distance matrix of size $N$ is the squared distance among its representing levels.

$$D = \Delta^2 \left\{ (i-j)^2 \right\}_{i,j=0}^{N-1} =$$

$$= \Delta^2 \cdot \begin{bmatrix} 0 & 1 & 4 & 9 & \cdots & (N-1)^2 \\ 1 & 0 & 1 & 4 & & \\ 4 & 1 & 0 & 1 & \ddots & \\ 9 & 4 & 1 & 0 & \ddots & 4 \\ \vdots & & \ddots & \ddots & \ddots & 1 \\ (N-1)^2 & & & 4 & 1 & 0 \end{bmatrix} \qquad \text{(A.1)}$$

Where $\Delta$ is the distance between two consecutive reconstruction levels.

Throughout we assume $N \geq 3$.

**Property 1**: $Rank(D) = 3$

Proof: Note that

$$\frac{1}{\Delta^2} D = \left\{ (i-j)^2 \right\}_{i,j=0}^{N-1} =$$

$$= \left\{ i^2 \right\}_{i,j=0}^{N-1} + \left\{ j^2 \right\}_{i,j=0}^{N-1} - 2\underline{\phi} \cdot \underline{\phi}^T \qquad \text{(A.2)}$$

where $\underline{\phi} = \begin{bmatrix} 0 & 1 & \cdots & N-1 \end{bmatrix}^T$.

Hence, $D$ is the sum of three matrices, each of Rank=1.

Using the Perron-Frobenius theorem [La85, Section 15.3] for positive matrices we show that one eigenvalue of $D$ is positive, two are negative, and the zero eigenvalue has a multiplicity of $N$-3.

**Definition**: A matrix $D$ is primitive if all elements of $D^k$ are strictly positive for some positive integer $k$.

**Perron-Frobenius theorem**: If a matrix $D$ is nonnegative and primitive, then

1. $D$ has a positive eigenvalue, $r$, with algebraic multiplicity 1.
2. The eigenvector associated with $r$ has positive entries.
3. For any other eigenvalue of $D$, $\lambda$, $r > |\lambda|$.

**Property 2**: All elements of $D^2$ are strictly positive, hence $D$ is primitive.

Proof: The general element of $D^2$ is:

$$\left[D^2\right]_{ij} = \Delta^4 \sum_{k=0}^{N-1} (i-k)^2 \cdot (k-j)^2 \tag{A.3}$$

Since the size of $D^2$ is greater than three ($N \geq 3$), there is at least one strictly positive expression (for which $k \neq i$ and $k \neq j$).

**Property 3**: $D$ has one positive eigenvalue and two negative eigenvalues for every size $N$. The sum of these three eigenvalues is zero.

Proof: The sum of all eigenvalues is zero since $trace\{D\} = 0$. Property 1 causes the number of nonzero eigenvalues to be three. Since $D$ is primitive, is has a positive eigenvalue (Perron-Frobenius eigenvalue). This eigenvalue is strictly greater than the absolute value of all other eigenvalues. Therefore, the remaining nonzero eigenvalues are negative.

<u>**Property 4**</u>: The nonzero eigenvalues are:

$$\alpha_0 = \Delta^2\left[\frac{1}{12}N\left(N^2-1\right)+N\cdot S_N\right] \quad \text{(positive)}$$

$$\alpha_1 = \Delta^2\left[\frac{1}{12}N\left(N^2-1\right)-N\cdot S_N\right]$$

$$\alpha_2 = -\Delta^2\frac{1}{6}N\left(N^2-1\right) \tag{A.4}$$

where $S_N = \sqrt{\dfrac{3N^4-10N^2+7}{240}}$ .

The corresponding eigenvectors:

$$\underline{u}_0 = \left[\left(i-\frac{N-1}{2}\right)^2\right]_{i=0}^{N-1} + S_N\cdot\underline{1}$$

$$\underline{u}_1 = \left[\left(i-\frac{N-1}{2}\right)^2\right]_{i=0}^{N-1} - S_N\cdot\underline{1}$$

$$\underline{u}_2 = [2i]_{i=0}^{N-1} - (N-1)\cdot\underline{1} \tag{A.5}$$

where $\underline{1} = \begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix}^T$.

<u>Proof</u>: The Proof can be done by a brute-force calculation.

For example:

$$\frac{1}{\Delta^2}[D\cdot\underline{u}_2]_i = \sum_{j=0}^{N-1}(i-j)^2[2j-N+1]=$$

$$=2\sum_{j=0}^{N-1}j^3+(1-4i-N)\sum_{j=0}^{N-1}j^2+\left(2i^2-2i+2iN\right)\sum_{j=0}^{N-1}j+i^2N-i^2N^2=$$

$$=\frac{1}{2}N^2(N-1)^2+\frac{1}{6}(1-4i-N)(N-1)N(2N-1)+$$

$$+\left(i^2-i+iN\right)(N-1)N+i^2N-i^2N^2=$$

$$=-\frac{1}{6}(N-1)N(N+1)\cdot[2i-N+1]=$$

$$=\frac{1}{\Delta^2}\alpha_2[\underline{u}_2]_i \tag{A.6}$$

# Appendix B

# נספח ב'

## האופטימליות של הקוד הבינארי הטבעי עבור קוונטייזר סקלרי אחיד, מקור אחיד ותחת הערוץ הבינארי הסימטרי

## The optimality of the Natural Binary Code Assignment for the Uniform Scalar Quantizer and a Uniform Source under the Binary Symmetric Channel

For the special case of a uniform scalar quantizer and a uniform source under the Binary Symmetric Channel, it is shown here that the Natural Binary Code offers the minimal Channel Distortion over all possible code assignments. This result is used in **Chapter** 5. The same result can be found in [Cr69],[Mc92], yet the optimization method as shown here is extended in Appendix C and Chapter 6 to give upper and lower bounds on general Vector Quantizers and general Memoryless Channels.

For the $L$-bit ($N = 2^L$ reconstruction levels) uniform scalar quantizer, we organize the $N$ reconstruction levels in a column vector:

$$\underline{\rho} = \begin{bmatrix} \rho_0 \\ \rho_1 \\ \vdots \\ \rho_i \\ \vdots \\ \rho_{N-1} \end{bmatrix} = \frac{2}{N} \cdot \begin{bmatrix} -N/2+1/2 \\ 1-N/2+1/2 \\ \vdots \\ i-N/2+1/2 \\ \vdots \\ N/2-1+1/2 \end{bmatrix} \tag{B.1}$$

We denote the channel probability matrix in the case of an $N$ levels quantizer by $Q_N$. The source probabilities matrix $P$ is defined as $P = diag([p_0 \quad p_1 \quad \cdots \quad p_{N-1}])$, where $p_i = \Pr\{\rho_i \text{ selected}\}$. Since the source is uniform $P = \frac{1}{N} \cdot I$, where $I$ is the unity matrix. The distortion measure of interest is the quadratic distance:

$$d(\rho_i, \rho_j) = (\rho_i - \rho_j)^2 \tag{B.2}$$

The channel distortion is (2.15):

$$D_c = trace\{P\pi Q_N \pi^T D\} =$$

$$\Leftarrow trace\{AB\} = trace\{BA\}, \quad P = \frac{1}{N}\cdot I$$

$$= \frac{1}{N} trace\{Q_N \pi^T D\pi\} =$$

$$\Leftarrow trace\{AB\} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(A)_{ij}\left(B^T\right)_{ij}$$

$$= \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(Q_N)_{ij}\left[\left(\pi\underline{\rho}\right)_i - \left(\pi\underline{\rho}\right)_j\right]^2 =$$

$$= \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(Q_N)_{ij}\left[z_i - z_j\right]^2 =$$

$$= \frac{1}{N}\sum_{i=0}^{N-1} z_i^2 \sum_{j=0}^{N-1}(Q_N)_{ij} + \frac{1}{N}\sum_{j=0}^{N-1} z_j^2 \sum_{i=0}^{N-1}(Q_N)_{ij} - \frac{2}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(Q_N)_{ij}\cdot z_i \cdot z_j =$$

$$= \frac{2}{N}\left(\underline{z}^T \underline{z} - \underline{z}^T Q_N \underline{z}\right) \tag{B.3}$$

where $\underline{z} = \pi\underline{\rho}$ is a permutation of the reconstruction levels.

In order to find lower and upper bounds for (B.3) over all possible assignments, it is suggested in [Mc92] to allow the entries of $\underline{z}$ to be continuous, while preserving the first two moments. The distortion measure $d(\rho_i, \rho_j)$ is not affected by a constant offset addition to the entries of $\underline{\rho}$ (and $\underline{z}$). This allows us to choose the sum of the entries in $\underline{z}$ to be zero.

The following optimization problem can be stated:

$$\begin{aligned}
\min_{\underline{z}} / \max_{\underline{z}} \quad & f(\underline{z}) = \underline{z}^T \underline{z} - \underline{z}^T Q_N \underline{z} \\
\text{s.t.} \quad & \underline{1}^T \underline{z} = 0 \\
& \underline{z}^T \underline{z} = \underline{\rho}^T \underline{\rho} = k^2 \\
& \Downarrow \\
\max_{\underline{z}} / \min_{\underline{z}} \quad & \tilde{f}(\underline{z}) = \underline{z}^T Q_N \underline{z} \\
\text{s.t.} \quad & \underline{1}^T \underline{z} = 0 \\
& \underline{z}^T \underline{z} = \underline{\rho}^T \underline{\rho} = k^2
\end{aligned} \tag{B.4}$$

where $\underline{1}$ is a $N$-dimensional vector with unity entries, and $k^2$ is the variance of the reconstruction levels, $k^2 = \dfrac{(N+1)(N-1)}{3N}$

Using Lagrange multipliers:

$$L(\underline{z},\mu_1,\mu_2) = \underline{z}^T Q_N \underline{z} + \mu_1\left(\underline{z}^T \underline{z} - k^2\right) + \mu_2\left(\underline{1}^T \underline{z}\right) \tag{B.5}$$

and the derivative of $L(\underline{z},\mu_1,\mu_2)$ w.r.t. $\underline{z}$:

$$\begin{aligned} \frac{\partial}{\partial \underline{z}} L(\underline{z},\mu_1,\mu_2) &= 2\underline{z}^T Q_N + 2\mu_1 \underline{z}^T + \mu_2 \underline{1}^T = \\ &= 2\underline{z}^T (Q_N + \mu_1 I) + \mu_2 \underline{1}^T \end{aligned} \tag{B.6}$$

one may obtain the following first order conditions:

$$\begin{aligned} &a. \quad (Q_N + \mu_1 \cdot I)\cdot \underline{z} = -\frac{\mu_2}{2}\cdot \underline{1} \\ &b. \qquad\quad 1^T \underline{z} = 0 \\ &c. \qquad\quad \underline{z}^T \underline{z} = k^2 \end{aligned} \tag{B.7}$$

Multiplying (B.7a) by $\underline{z}^T$, we have:

$$\begin{aligned} \underline{z}^T \cdot (Q_N + \mu_1 \cdot I)\cdot \underline{z} &= -\frac{\mu_2}{2}\cdot \underbrace{\underline{z}^T \cdot \underline{1} \qquad = 0}_{(B.7b)} \\ \therefore \quad \underline{z}^T Q_N \underline{z} &= -\mu_1 k^2 \\ \therefore \quad \mu_1 &= -\frac{1}{k^2}\underline{z}^T Q_N \underline{z} \end{aligned} \tag{B.8}$$

We define a normalized vector $\underline{w} = \dfrac{1}{k}\cdot \underline{z}$ and substitute (B.8) into (B.7a):

$$\left(Q_N - \underline{w}^T Q_N \underline{w}\cdot I\right)\cdot \underline{w} = -\frac{\mu_2}{2k}\cdot \underline{1} \tag{B.9}$$

For the Binary Symmetric Channel, $Q_N$ is symmetric and therefore has a unitary diagonalization:

$$Q_N = V \cdot \Lambda \cdot V^T \qquad V \cdot V^T = I \tag{B.10}$$

Since $Q_N$ represents probabilities, the sum of any of its rows is one, so $\underline{1}$ is an eigenvector of $Q_N$.

$$Q_N \cdot \underline{1} = \underline{1} \tag{B.11}$$

We may assume, therefore, that the first column of $V$ is $\frac{1}{\sqrt{N}} \cdot \underline{1}$.

In order to simplify (B.9), we use now the following unitary transform $\underline{\xi} = V^T \underline{w}$. This vector has the following properties:

$$\begin{aligned} &\text{a. } \underline{\xi}^T \underline{\xi} = \underline{w}^T V V^T \underline{w} = \underline{w}^T \underline{w} = 1 \\ &\text{b. } \xi_0 = \frac{1}{\sqrt{N}} \cdot \underline{1}^T \underline{w} = 0 \end{aligned} \tag{B.12}$$

Where $\xi_0$ is the first entry of $\underline{\xi}$.

The first order conditions in equation (B.9) are:

$$\begin{aligned} &\left( \overbrace{V \Lambda V^T}^{Q_N} - \underline{w}^T \overbrace{V \Lambda V^T}^{Q_N} \underline{w} \cdot I \right) \cdot \underline{w} = -\frac{\mu_2}{2k} \cdot \underline{1} \\ \therefore \quad &\left( V \Lambda V^T - \underline{\xi}^T \Lambda \underline{\xi} \overbrace{V V^T}^{I} \right) \cdot \underline{w} = -\frac{\mu_2}{2k} \cdot \underline{1} \\ \therefore \quad &V \left( \Lambda - \underline{\xi}^T \Lambda \underline{\xi} \right) \cdot V^T \underline{w} = -\frac{\mu_2}{2k} \cdot \underline{1} \\ \therefore \quad &\left( \Lambda - \underline{\xi}^T \Lambda \underline{\xi} \right) \cdot \underline{\xi} = -\frac{\mu_2}{2k} \cdot V^T \underline{1} \end{aligned} \tag{B.13}$$

The first row of $V^T$ is $\frac{1}{\sqrt{N}}\cdot\underline{1}^T$, and the rest of its rows are orthogonal w.r.t. $\underline{1}^T$.

Therefore:

$$\therefore \quad \overbrace{\left(\Lambda-\underline{\xi}^T\Lambda\underline{\xi}\cdot I\right)}^{\text{diagonal matrix}}\cdot\underline{\xi}=-\frac{\mu_2}{2k}\cdot V^T\cdot\underline{1}=\begin{bmatrix}-\frac{\mu_2\sqrt{N}}{2k}\\ 0\\ \vdots\\ 0\end{bmatrix} \tag{B.14}$$

Observing that the first line of the diagonal matrix in (B.14) (the first eigenvalue of $Q_N$ is $\lambda_0=1$),

$$\left[\lambda_0-\left(\underline{\xi}^T\Lambda\underline{\xi}\right)\right]\cdot\xi_0=-\frac{\mu_2\sqrt{N}}{2k} \tag{B.15}$$

and since $\xi_0$ is zero (B.12b), $\mu_2$ is also zero. The Lagrange multiplier $\mu_2$ represents the "sensitivity" of the optimization problem with respect to changes in the corresponding constraint. Namely, replacing (B.7b) with $\underline{1}^T\underline{z}=c$, for a sufficiently small range of $c$ near zero, the optimization problem will have a solution point $\underline{z}(c)$ near the original point $\underline{z}(0)$. For each solution there is a corresponding value $\tilde{f}(\underline{z}(c))$ of (B.4), and this value can be regarded as a scalar function of $c$. It is shown in [Lu84, Section 10.7] that the derivative of this function w.r.t. $c$ is $\mu_2$.

$$\left.\frac{d}{dc}\tilde{f}(\underline{z}(c))\right|_{c=0}=\mu_2 \tag{B.16}$$

We have just found that our optimization problem is not sensitive ($\mu_2=0$) to an addition of constants to all entries of $\underline{z}$, as was mentioned in (B.4).

The remaining lines of the diagonal matrix in (B.14)

$$\left[\lambda_i-\left(\underline{\xi}^T\Lambda\underline{\xi}\right)\right]\cdot\xi_i=0,\quad i=1,2,\ldots,N-1 \tag{B.17}$$

Where $\lambda_i$ are the eigenvalues of $Q_N$. Note that some entries of $\underline{\xi}$ must be non-zero because of (B.12a).

Assume some eigenvalue of $Q_N$ - $\lambda$, has a multiplicity of $m$:

$$\lambda_l = \lambda_{l+1} = \cdots = \lambda_{l+m-1} = \lambda \tag{B.18}$$

A solution corresponding to this eigenvalue is as follows.

Equation (B.17) forces in this case $\lambda = \underline{\xi}^T \Lambda \underline{\xi}$. Moreover, only entries of $\underline{\xi}$ corresponding to those eigenvalues, $\xi_i \quad i = l, l+1, \ldots, l+m-1$ can be nonzero.

The original vector $\underline{z}$,

$$\underline{z} = k \cdot \underline{w} = k \cdot V \underline{\xi} \tag{B.19}$$

is a linear combination of the columns of $V$ (eigenvectors of $Q_N$). Since only $\xi_i$, $i = l, l+1, \ldots, l+m-1$, are nonzero, $\underline{z}$ is a linear combination of the eigenvectors corresponding with the eigenvalue $\lambda$. This vector $\underline{z}$ should also comply with the variance constraint in (B.4).

Substitute (B.19) into (B.4)

$$\begin{aligned} f(\underline{z}) &= k^2 - k^2 \underline{\xi}^T V^T \cdot V \Lambda V^T \cdot V \underline{\xi} = \\ &= k^2\left(1 - \underline{\xi}^T \Lambda \underline{\xi}\right) = k^2(1 - \lambda) \end{aligned} \tag{B.20}$$

One can see now that the minimal eigenvalue maximizes $f(\underline{z})$, while the maximal eigenvalue (not $\lambda_0 = 1$) minimizes it.

We turn now to calculate the eigenvalues and eigenvectors of $Q_N$. The calculation is done recursively in the number of bits - $L$.

For the single bit (two levels) uniform quantizer:

$$\begin{aligned} &a. \quad Q_2\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1-q & q \\ q & 1-q \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &b. \quad Q_2\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1-q & q \\ q & 1-q \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} = (1-2q)\begin{bmatrix} 1 \\ -1 \end{bmatrix} \end{aligned} \tag{B.21}$$

Assume that $\underline{v}$ is an eigenvector of $Q_N$ ,i.e., $Q_N \underline{v} = \alpha \underline{v}$, it is possible to construct two eigenvectors for a quantizer with twice the number of level ($L+1$ bits):

a. *duplication*

$$Q_{2N}\begin{bmatrix} \underline{v} \\ \underline{v} \end{bmatrix} = \begin{bmatrix} (1-q)Q_N & qQ_N \\ qQ_N & (1-q)Q_N \end{bmatrix} \cdot \begin{bmatrix} \underline{v} \\ \underline{v} \end{bmatrix} =$$

$$= \begin{bmatrix} Q_N \underline{v} \\ Q_N \underline{v} \end{bmatrix} = \alpha \begin{bmatrix} \underline{v} \\ \underline{v} \end{bmatrix}$$

b. *negative – duplication*

$$Q_{2N}\begin{bmatrix} \underline{v} \\ -\underline{v} \end{bmatrix} = \begin{bmatrix} (1-q)Q_N & qQ_N \\ qQ_N & (1-q)Q_N \end{bmatrix} \cdot \begin{bmatrix} \underline{v} \\ -\underline{v} \end{bmatrix} =$$

$$= \begin{bmatrix} (1-2q)Q_N \underline{v} \\ -(1-2q)Q_N \underline{v} \end{bmatrix} = (1-2q)\alpha \begin{bmatrix} \underline{v} \\ -\underline{v} \end{bmatrix} \tag{B.22}$$

The eigenvectors are constructed by either *duplication* of a smaller size eigenvector, corresponding to the same eigenvalue, or by *negative-duplication* corresponding to the eigenvalue multiplied by $1-2q$. Observing the recursive relations, one may verify that the eigenvectors matrix $V$ is a Walsh-Hadamard matrix [Cl85, Section 7.3].

Using these recursive relations the first three eigenvalue, eigenvectors sets are:

$$L=1 \qquad 1, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad 1-2q, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{B.23a}$$

$$L=2 \qquad 1, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad 1-2q, \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \qquad 1-2q, \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \qquad (1-2q)^2, \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \tag{B.23b}$$

$$L=3 \quad 1,\begin{bmatrix}1\\1\\1\\1\\1\\1\\1\\1\end{bmatrix} \quad 1-2q,\begin{bmatrix}1\\1\\-1\\-1\\1\\1\\-1\\-1\end{bmatrix} \quad 1-2q,\begin{bmatrix}1\\-1\\1\\-1\\1\\-1\\1\\-1\end{bmatrix} \quad (1-2q)^2,\begin{bmatrix}1\\-1\\-1\\1\\1\\-1\\-1\\1\end{bmatrix}$$

$$1-2q,\begin{bmatrix}1\\1\\1\\1\\-1\\-1\\-1\\-1\end{bmatrix} \quad (1-2q)^2,\begin{bmatrix}1\\1\\-1\\-1\\-1\\-1\\1\\1\end{bmatrix} \quad (1-2q)^2,\begin{bmatrix}1\\-1\\1\\-1\\-1\\1\\-1\\1\end{bmatrix} \quad (1-2q)^3,\begin{bmatrix}1\\-1\\-1\\1\\-1\\1\\1\\-1\end{bmatrix} \tag{B.23c}$$

In general, there are $L+1$ distinct eigenvalues, $(1-2q)^m \quad m=0,1,\ldots,L$, each with multiplicity $\binom{L}{m}$.

Using the known eigenvalues, we can now set bounds on $f(\underline{z})$ (B.20).

$$k^2 \cdot 2q \le f(\underline{z}) \le k^2\left[1-(1-2q)^L\right]$$

$$\therefore \frac{2(N+1)(N-1)}{3N^2}\cdot 2q \le D_c \le \frac{2(N+1)(N-1)}{3N^2}\left[1-(1-2q)^L\right] \tag{B.24}$$

For low BER, the ratio between the upper and lower bounds is:

$$\lim_{q\to 0}\frac{\left[1-(1-2q)^L\right]}{2q}=L \tag{B.25}$$

Up to this point we know that a vector $\underline{z}$ which is an eigenvector of $Q_N$, corresponding to the eigenvalue $1-2q$, minimizes the target function $f(\underline{z})$. An eigenvector corresponding to the eigenvalue $(1-2q)^L$ maximizes the target function.

The last statement to be verified is that the Natural Binary Code indeed results in an eigenvector corresponding to the eigenvalue $1-2q$.

For an $L$-bits quantizer, the eigenvectors of $\underline{Q}_N$ as constructed in (B.21) and (B.22), corresponding to the eigenvalue $1-2q$, are constructed by $L-1$ duplications and a single negative-duplication.

For instance when the first duplication is negative we have,

$$\underline{v}_1 = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & -1 & \cdots & \cdots & 1 & -1 \end{bmatrix}^T \tag{B.26}$$

Note that the elements of $\underline{v}_1$ are $(-1)^i$, $i = 0,1,\cdots,N-1$.

When the second duplication is negative, the elements of $\underline{v}_2$ are $(-1)^{\lfloor i/2 \rfloor}$, $i = 0,1,\cdots,N-1$, where $\lfloor i/2 \rfloor$ stands for integer division .

$$\underline{v}_2 = \begin{bmatrix} 1 & 1 & -1 & -1 & 1 & 1 & \cdots & \cdots & -1 & -1 \end{bmatrix}^T \tag{B.27}$$

The last eigenvector corresponding to $1-2q$ is constructed such that the last duplication is negative, the first $N/2$ elements are +1, and the rest are -1.

$$\underline{v}_L = \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 & -1 & -1 & \cdots & -1 & -1 \end{bmatrix}^T \tag{B.28}$$

Note the similarity between the binary enumeration system and the entries of the $L$ eigenvectors - $\underline{v}_1$ - "*Least Significant Bit*", $\underline{v}_2, \underline{v}_3, \cdots, \underline{v}_{L-1}$ and $\underline{v}_L$ - "*Most Significant Bit*".

For example, assigning the entries of $\underline{\xi}$ to be:

$$\begin{aligned} &\xi_1 = \frac{-1}{N}, \xi_2 = \frac{-2}{N}, \xi_3 = \frac{-4}{N}, \cdots \xi_L = \frac{-2^{L-1}}{N} = -\frac{1}{2} \\ &\text{and } \xi_0 = 0, \quad \xi_{L+1} = \cdots = \xi_{N-1} = 0 \end{aligned} \tag{B.29}$$

results in $\underline{z} = \underline{\rho}$, since:

$$\underline{z} = \sum_{i=1}^{L} \xi_i \cdot \underline{v}_i = \frac{1}{N}[-1 \quad 1 \quad -1 \quad 1 \quad -1 \quad 1 \quad \cdots \quad \cdots \quad -1 \quad 1]^T +$$
$$+\frac{2}{N}[-1 \quad -1 \quad 1 \quad 1 \quad -1 \quad -1 \quad \cdots \quad \cdots \quad 1 \quad 1]^T +$$
$$\vdots$$
$$+\frac{1}{2}[-1 \quad -1 \quad \ldots \quad -1 \quad -1 \quad 1 \quad 1 \quad \cdots \quad 1 \quad 1]^T = \underline{\rho} \tag{B.30}$$

These are the reconstruction levels of the Natural Binary Code assignment $\pi = I$.
For the same reasons, the positive values:

$$\xi_1 = \frac{1}{N}, \xi_2 = \frac{2}{N}, \xi_3 = \frac{4}{N}, \cdots \xi_L = \frac{2^{L-1}}{N} = \frac{1}{2} \tag{B.31}$$
$$\text{and } \xi_0 = 0, \quad \xi_{L+1} = \cdots = \xi_{N-1} = 0$$

are the linear coefficients of $\underline{z}$ corresponding to the "down enumerating" Binary Code

$$\pi = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ & & \ddots & \vdots \\ 0 & 1 & & 0 \\ 1 & 0 & & 0 \end{bmatrix}.$$

Moreover, assigning $\xi_1, \xi_2, \ldots, \xi_L$ with every permutation of the $L$ values (each can be chosen positive or negative) $\pm\frac{1}{N}, \pm\frac{2}{N}, \pm\frac{4}{N}, \cdots, \pm\frac{2^{L-1}}{N} = \frac{1}{2}$ will result in a feasible permutation of $\underline{\rho}$ (B.1). Therefore, we can construct in this way $\underbrace{2^L}_{\text{positive - negative}} \cdot \underbrace{L!}_{\text{permutation}}$ optimal assignments, all with the same value of channel distortion $D_c$.

Worst assignment cannot be constructed in this way, since only one eigenvector exits for the eigenvalue $(1-2q)^L$.

◊

# Appendix C

# נספח ג'

## חסמים על עיוות הערוץ מעל כל השיבוצים האפשריים עבור ערוצים סימטריים חסרי-זיכרון

## Bounds on the Channel Distortion for Symmetric Memoryless Channels over all possible assignments

In this appendix we introduce lower and upper bounds on the channel distortion under Symmetric Memoryless Channels over all possible assignments. These bounds are used in chapter 6, where numerical results are presented. We assume knowledge of the source probabilities matrix $P$, and the VQ codevectors distance matrix $D$.

**Phase 1**: At first we combine the matrices $D$ and $P$ into a single symmetric matrix. The channel distortion is (2.15):

$$
\begin{aligned}
D_c &= trace\{P\pi Q_N \pi^T D\} = \\
&\qquad\qquad \Leftarrow trace\{AB\} = trace\{BA\} \\
&= \frac{1}{2}\left[trace\{\pi Q_N \pi^T DP\} + trace\{DP\pi Q_N \pi^T\}\right] = \\
&\qquad\qquad \Leftarrow trace(A) = trace(A^T) \\
&= \frac{1}{2} trace\{\pi Q_N \pi^T (DP + P^T D^T)\} = \\
&\qquad\qquad \Leftarrow trace\{AB\} = trace\{BA\} \\
&= \frac{1}{2} trace\left\{ Q_N \pi^T \underbrace{(DP + P^T D^T)}_{\hat{D}} \pi \right\}
\end{aligned}
\tag{C.1}
$$

where $\pi$ is an unknown permutation matrix.

The bounding technique is based on eigenvalues arguments. Instead of optimizing over the (discrete) family of matrices covering all possible assignments $\pi Q_N \pi^T$, we optimize over a wider (continuous) family. A fundamental step in this optimization procedure is that the matrix $\hat{D} = DP + P^T D^T$ is replaced by another symmetric matrix $\tilde{D}$, having $\underline{1} = [1 \ \ 1 \ \ \cdots \ \ 1]^T$ as an eigenvector. This goal is achieved in phase 2.

**Phase 2**: We modify (C.1) so that the symmetric matrix $\hat{D}$ is replaced by another symmetric matrix $\tilde{D}$, having the property $\tilde{D}\cdot\underline{1}=\beta\underline{1}$, where $\underline{1}=[1 \quad 1 \quad \cdots \quad 1]^T$.

In order to achieve this property we add to the symmetric matrix $\hat{D}$ matrices of the following structure.

$$C_i = \begin{array}{c} \begin{bmatrix} 0 & & 0 & 1 & 0 & & & 0 \\ 0 & & 0 & 1 & 0 & & & 0 \\ 0 & & 0 & 1 & 0 & & & 0 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & & 0 & 1 & 0 & & & 0 \\ 0 & & 0 & 1 & 0 & & & 0 \\ 0 & & 0 & 1 & 0 & & & 0 \end{bmatrix} \\ \uparrow \; i-\text{th column} \end{array} \tag{C.2}$$

Recalling that $Q_N$ represents probabilities, the sum of any of its rows is one, so the vector $\underline{1}$ an eigenvector of $Q_N$.

$$\begin{aligned} &Q_N\cdot\underline{1}=\underline{1} \\ &\therefore Q_N\cdot C_i = C_i \end{aligned} \tag{C.3}$$

We show now that while preserving the symmetry of the matrix $\hat{D}$, we may add "*Cross Structured*" matrices $\alpha(C_i+C_i^T)$, where $\alpha$ is a scalar. This changes the expression $\frac{1}{2}trace\{Q_N\pi^T\hat{D}\pi\}$ in (C.1) just by the addition of $\alpha$.

**Property 1**: Adding a scalar multiple of $C_i$ to the matrix $\hat{D}=PD+DP$:

$$\begin{aligned} &trace\{Q_N\pi^T(\hat{D}+\alpha C_i)\pi\} = \\ &= trace\{Q_N\pi^T\hat{D}\pi\}+\alpha\cdot trace\{Q_N\pi^T C_i\pi\} = \\ &= trace\{Q_N\pi^T\hat{D}\pi\}+\alpha\cdot trace\{Q_N C_i\pi\} = \\ &= trace\{Q_N\pi^T\hat{D}\pi\}+\alpha\cdot trace\{\pi Q_N C_i\} = \\ &= trace\{Q_N\pi^T\hat{D}\pi\}+\alpha\cdot trace\{\pi C_i\} = \\ &= trace\{Q_N\pi^T\hat{D}\pi\}+\alpha \end{aligned} \tag{C.4}$$

Similarly, for $C_i^T$:

$$\begin{aligned}
&trace\left\{Q_N\pi^T\left(\hat{D}+\alpha C_i^T\right)\pi\right\}= \\
&=trace\left\{Q_N\pi^T\hat{D}\pi\right\}+\alpha\cdot trace\left\{Q_N\pi^T C_i^T\pi\right\}= \\
&=trace\left\{Q_N\pi^T\hat{D}\pi\right\}+\alpha\cdot trace\left\{Q_N\pi^T C_i^T\right\}= \\
&=trace\left\{Q_N\pi^T\hat{D}\pi\right\}+\alpha\cdot trace\left\{C_i^T Q_N\pi^T\right\}= \\
&=trace\left\{Q_N\pi^T\hat{D}\pi\right\}+\alpha\cdot trace\left\{C_i^T\pi^T\right\}= \\
&=trace\left\{Q_N\pi^T\hat{D}\pi\right\}+\alpha
\end{aligned} \tag{C.5}$$

Applying both (C.5) and (C.6):

$$\begin{aligned}
&\frac{1}{2}trace\left\{Q_N\pi^T\left(\hat{D}+\alpha C_i^T+\alpha C_i\right)\pi\right\}= \\
&\quad=\frac{1}{2}trace\left\{Q_N\pi^T\hat{D}\pi\right\}+\alpha
\end{aligned} \tag{C.6}$$

**<u>Algorithm:</u>** In order to achieve the desired property $\tilde{D}\cdot\underline{1}=\beta\underline{1}$, for some $\beta$, all rows of $\tilde{D}$ must have the same sum of elements. Let us examine the effect of adding a "*Cross Structured*" matrix $\alpha\left(C_i+C_i^T\right)$ to a any matrix $M$ of size $N\times N$. The sum of all rows but the $i$-th row is increased by $\alpha$. The sum of the $i$-th row is increased by $(N+1)\cdot\alpha$. Throughout the algorithm, a variable $S$ is needed to store the sum of all "$\alpha$" constant added to the r.h.s. of (C.1).

*Pseudo-Code Listing for phase 2*

Initialization: a. Set the matrix: $\tilde{D} \leftarrow \hat{D} = DP + P^T D^T$.

b. Set the sum of additive constants to zero: $S \leftarrow 0$.

Step 1: Calculate the sum of all rows.

Denote the sum of the $i$-th row by $S_i = \sum_{j=0}^{N-1} \left(\tilde{D}\right)_{ij}$.

Step 2: Search all rows for the maximal sum of elements.

Assume that the row with the maximal sum is labeled $k$.

Step 3: For each row $i \neq k$:

a. Add the "*Cross Structured*" matrix $\tilde{D} \leftarrow \tilde{D} + \frac{1}{N}(S_k - S_i)\left(C_i + C_i^T\right)$.

b. Update $S \leftarrow S + \frac{1}{N}(S_k - S_i)$.

After execution of step 3, all rows have the same sum of elements.

By adding $N$-1 "*Cross Structured*" matrices we get a symmetric matrix where all rows have the same sum of elements, resulting in the matrix $\tilde{D}$, with the desired property $\tilde{D} \cdot \underline{1} = \omega_0 \underline{1}$.

**Phase 3**: The channel distortion is now:

$$D_c = \frac{1}{2} trace\{Q_N \pi^T \tilde{D} \pi\} - S \tag{C.7}$$

The symmetric channel probability matrix $Q_N$ has a unitary diagonalization:

$$Q_N = V \cdot \Lambda \cdot V^T \qquad V \cdot V^T = I \tag{C.8}$$

We shall use the following theorem from [La85, Section 15.7].

Theorem: The Perron-Frobenius eigenvalue of a nonnegative symmetric matrix $M$ with the property $M \cdot \underline{1} = \beta \underline{1}$ is $\beta$.

We arrange the eigenvalues in $\Lambda$ to be in decreasing order. Therefore, using the above theorem , the first column of $V$ is $\frac{1}{\sqrt{N}} \cdot \underline{1}$ and $\lambda_0 = 1$.

Substituting (C.8) into (C.7)

$$\begin{aligned} D_c &= \frac{1}{2} trace\{V \Lambda V^T \pi^T \tilde{D} \pi\} - S = \\ &= \frac{1}{2} trace\{\pi V \Lambda V^T \pi^T \tilde{D}\} - S = \end{aligned} \tag{C.9}$$

The matrix $\pi V$ is unitary for every permutation-matrix $\pi$, since $(\pi V)(\pi V)^T = \pi V V^T \pi^T = \pi \pi^T = I$. Moreover, for the first eigenvalue of $Q_N$, $\lambda_0 = 1$ (C.3), permutation does not affect its corresponding eigenvector: $\pi \underline{v}_0 = \frac{1}{\sqrt{N}} \pi \underline{1} = \frac{1}{\sqrt{N}} \underline{1}$.

In order to obtain the desired bounds we broaden the set of possible unitary matrices. Instead of optimizing $D_c$ over all possible permutations (discrete), we bound the channel distortion by optimizing over all unitary matrices $U$, having $\underline{1}$ as the eigenvector corresponding to $\lambda_0 = 1$ (continuous).

The following optimization problem is to be solved:

$$\min_U / \max_U \left( trace\{\Lambda U^T \tilde{D} U\} \right) \tag{C.10}$$

where $UU^T = I$ and $\underline{u}_0 = \frac{1}{\sqrt{N}}\underline{1}$.

Note that the symmetric matrix $\tilde{D}$ has also a unitary diagonalization $\tilde{D} = Z \cdot \Omega \cdot Z^T$, with the eigenvalues $\omega_i$ and corresponding eigenvectors $\underline{z}_i$. Moreover, $\tilde{D}$ was constructed to have the eigenvector $\underline{z}_0 = \frac{1}{\sqrt{N}}\underline{1}$ with a corresponding eigenvalue $\omega_0$. The eigenvalue $\omega_0$ is the Perron-Frobenius eigenvalue of $\tilde{D}$. All other eigenvalues are arranged in decreasing order.

The matrix $\Theta = U^T Z$ is unitary, since $\Theta^T \Theta = Z^T U U^T Z = Z^T Z = I$. Moreover, the first column of $\Theta$ is $\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, since both $U$ and $Z$ have orthonormal columns, and have $\underline{u}_0 = \underline{z}_0 = \frac{1}{\sqrt{N}}\underline{1}$.

Substitute $\tilde{D}$ into the optimization problem (C.10):

$$\min_U / \max_U \left( trace\{\Lambda U^T Z \cdot \Omega \cdot Z^T U\} \right)$$

$$\therefore \quad \min_\Theta / \max_\Theta \left( trace\{\Lambda \Theta \cdot \Omega \cdot \Theta^T\} \right)$$

$$\therefore \quad \min_\Theta / \max_\Theta \left( \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \lambda_i \omega_j \theta_{ij}^2 \right) \tag{C.11}$$

where $\Theta\Theta^T = I$ and $\underline{\theta}_0 = [1 \quad 0 \quad \cdots \quad 0]^T$.

The sum of squares of the elements in each row and column of a unitary matrix is equal to 1.

We relax the constraints in (C.11) and state the following problem:

$$\begin{aligned} & \min_{\Theta} / \max_{\Theta} \left( \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \lambda_i \omega_j \theta_{ij}^2 \right) \\ s.t. \quad & \sum_{i=0}^{N-1} \theta_{ij}^2 = 1 \quad j = 0,1,\dots,N-1 \\ & \sum_{j=0}^{N-1} \theta_{ij}^2 = 1 \quad i = 0,1,\dots,N-1 \end{aligned} \tag{C.12}$$

This is a standard Assignment problem [Ta87, Section 6.3]. An optimal solution for an Assignment problem is a permutation matrix $\Theta_{opt}$ that has a single 1 in each row and column, while the rest of the matrix are zero. Such matrix is also unitary, thus solving (C.11), as long as the first column is $\underline{\theta}_0 = [1 \quad 0 \quad \cdots \quad 0]^T$.

Observing the target function in (C.11) $\sum_{i=0}^{N-1} \lambda_i \sum_{j=0}^{N-1} \omega_j \theta_{ij}^2$ , we see that the permutation matrix $\Theta_{opt}$ does a one-to-one matching between the eigenvalues $\lambda_i$ and $\omega_i$ (always matching $\lambda_0$ and $\omega_0$, since $\underline{\theta}_0 = [1 \quad 0 \quad \cdots \quad 0]^T$).

Since both $\lambda_i$ and $\omega_i$ were sorted in decreasing order, the highest (lowest) possible value is obtained by matching the remaining eigenvalues in the same (reversed) order.[1]

---

[1] It is easy to verify this for series of two elements: $a_1 \geq a_2$ and $b_1 \geq b_2 \Rightarrow$ $(a_1 b_1 + a_2 b_2) - (a_1 b_2 + a_2 b_1) = (a_1 - a_2)(b_1 - b_2) \geq 0$.

Now assume $a_0 \geq a_1 \geq \cdots \geq a_{N-1}$ and $b_0 \geq b_1 \geq \cdots \geq b_{N-1}$, and we would like to maximize $S(\Pi) = \sum_{i=0}^{N-1} a_i b_{\Pi(i)}$ where $\Pi(i)$ is a permutation of the index $i$. If $b_{\Pi(i)} < b_{\Pi(i+1)}$, then $S(\Pi)$ would be larger if we switch them.

We may continue looking for such disordered constant $b_{\Pi(i)}$, while sorting them like in a **Bubble-Sort** algorithm. This sort results in $\Pi(i) = i$ .

Same argument can be apply to check that minimum is achieved by $\Pi(i) = N - 1 - i$.

In conclusion, the minimum and maximum of the optimization problem are:

$$\lambda_0\omega_0 + \sum_{i=1}^{N-1} \lambda_i \cdot \omega_{N-i} \le trace\{\Lambda U^T \tilde{D} U\} \le \lambda_0\omega_0 + \sum_{i=1}^{N-1} \lambda_i \cdot \omega_i \qquad \text{(C.13)}$$

The upper and lower bounds are, therefore, obtained by the following algorithm:

***Lower and Upper Bounds Pseudo-Code Listing***

Note: First three step are similar to phase 2 algorithm.

Initialization: a. Set the matrix: $\tilde{D} \leftarrow \hat{D} = DP + P^T D^T$.

b. Set the sum of additive constants to zero: $S \leftarrow 0$.

Step 1: Calculate the sum of all rows.

Denote the sum of the $i$-th row by $S_i = \sum_{j=0}^{N-1} (\tilde{D})_{ij}$.

Step 2: Search all rows for the maximal sum of elements.

Assume that the row with the maximal sum is labeled $k$.

Step 3: For each row $i \ne k$:

a. Add the "*Cross Structured*" matrix $\tilde{D} \leftarrow \tilde{D} + \frac{1}{N}(S_k - S_i)(C_i + C_i^T)$.

b. Update $S \leftarrow S + \frac{1}{N}(S_k - S_i)$.

Step 4: Calculate the eigenvalues of the Channel Matrix $Q$, and sort them in decreasing order $\lambda_0 \ge \lambda_1 \ge \cdots \ge \lambda_{N-1}$. For the Binary Symmetric Channel use App. B.

Remark: $\lambda_0 = 1$ with a corresponding eigenvector $\underline{1}$.

*__Lower and Upper Bounds Pseudo-Code Listing (cont.)__*

Step 5: Calculate the eigenvalues of $\tilde{D}$, and sort them in decreasing order $\omega_0 \geq \omega_1 \geq \cdots \geq \omega_{N-1}$.

Remark: $\omega_0$ has a corresponding eigenvector $\underline{1}$.

Step 6: Calculate the Upper and lower bounds from:

$$\lambda_0\omega_0 + \sum_{i=1}^{N-1} \lambda_i \cdot \omega_{N-i} \leq trace\{\Lambda U^T \tilde{D} U\} \leq \lambda_0\omega_0 + \sum_{i=1}^{N-1} \lambda_i \cdot \omega_i$$

**Remark 1**: A degree of freedom exists while constructing $\tilde{D}$. A scalar multiple of the symmetric "*All ones*" matrix $F = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$ can be added to $\tilde{D}$, while preserving the symmetry property and keeping $\underline{1}$ as an eigenvector. Such an action does not affect the resulting lower and upper bounds.

Note the following properties, which arise from the orthogonality of $U$:

$$\begin{aligned} i = 0:&\quad \underline{u}_0^T F \underline{u}_0 = \frac{1}{N}\underline{1}^T F \underline{1} = N \\ i \neq 0:&\quad \underline{u}_i^T F \underline{u}_i = \underline{u}_i^T \begin{bmatrix} \underline{1}^T \\ \vdots \\ \underline{1}^T \end{bmatrix} \underline{u}_i = 0 \quad \Leftarrow \underline{u}_i \perp \underline{u}_0 \end{aligned} \qquad \text{(C.14)}$$

Is it now shown that the optimization problem in (C.10) is changed only by an additive constant, thus not affecting the optimal matrix $U$.

$$\sum_{i=0}^{N-1} \lambda_i \underline{u}_i^T \left(\tilde{D} + \gamma F\right) \underline{u}_i =$$

$$= \sum_{i=0}^{N-1} \lambda_i \underline{u}_i^T \tilde{D} \underline{u}_i + \gamma \sum_{i=0}^{N-1} \lambda_i \underline{u}_i^T F \underline{u}_i =$$

$$\Leftarrow \underline{u}_i^T F \underline{u}_i = 0 \text{ for } i \neq 0$$

$$= \sum_{i=0}^{N-1} \lambda_i \underline{u}_i^T \tilde{D} \underline{u}_i + \gamma \lambda_0 \underline{u}_0^{\prime T} F \underline{u}_0 =$$

$$\Leftarrow \underline{u}_0^T F \underline{u}_0 = N \text{ and } \lambda_0 = 1$$

$$= \sum_{i=0}^{N-1} \lambda_i \underline{u}_i^T \tilde{D} \underline{u}_i + \gamma N \tag{C.15}$$

The factor $\gamma N$ is also added to $S$ in step 3 of phase 2, thus it vanishes in (C.7).

<u>**Remark 2**</u>: The Rank of the matrix $\tilde{D}$ is greater than the Rank of $\left(DP + P^T D^T\right)$ by at most 2. Observing that the matrix $\tilde{D}$ can be written as:

$$\tilde{D} = DP + P^T D^T + \underbrace{\sum_{i=0}^{N-1} \alpha_i C_i}_{A} + \underbrace{\sum_{i=0}^{N-1} \alpha_i C_i^T}_{B} \tag{C.16}$$

Both matrices $A$ and $B$ have Rank=1.

<u>**Remark 3**</u>: For the case of a Uniform Scalar Quantizer and a Uniform Source, the lower and upper bounds obtained here coincide with those found in Appendix B.

<u>Proof</u>: In this case (see Appendix A):

$$\hat{D} = DP + P^T D^T = \frac{2}{N} D = \left(\frac{2}{N}\right)^3 \cdot \left\{(i-j)^2\right\}_{i,j=0}^{N-1} =$$
$$= \frac{8}{N^3} \cdot \left[\left\{i^2\right\}_{i,j=0}^{N-1} + \left\{j^2\right\}_{i,j=0}^{N-1} - 2\left\{i \cdot j\right\}_{i,j=0}^{N-1}\right] \tag{C.17}$$

In phase 2, the sum of all rows is made equal. The term $\left\{j^2\right\}_{i,j=0}^{N-1}$ is common to all rows, and can be neglected.

Step 1: The sum of all rows:

$$S_i = \frac{8}{N^3}\left[N \cdot i^2 - 2i\sum_{j=0}^{N-1} j\right] = \frac{8}{N^3}\left[N \cdot i^2 - N(N-1)\cdot i\right] =$$
$$= \frac{8}{N^2}\left[i^2 - (N-1)\cdot i\right] \quad i = 0,1,\ldots,N-1 \tag{C.18}$$

Step 2: The maximal sum is achieved at $i = 0$ and $i = N-1$.

We choose $k = 0$, hence $S_k = 0$.

Step 3: The sum of additive constants:

$$S = -\frac{1}{N}\sum_{i=0}^{N-1} S_i = -\frac{8}{N^3}\left(\sum_{i=0}^{N-1} i^2 - (N-1)\sum_{i=0}^{N-1} i\right) =$$
$$= -\frac{8}{N^3}\left(\frac{(N-1)N(2N-1)}{6} - (N-1)\frac{(N-1)N}{2}\right) =$$
$$= \frac{4(N-1)(N-2)}{3N^2} \tag{C.19}$$

The matrix $\tilde{D}$:

$$\tilde{D} = \frac{8}{N^3}\cdot\left[\{i^2\}_{i,j=0}^{N-1} + \{j^2\}_{i,j=0}^{N-1} - 2\{i\cdot j\}_{i,j=0}^{N-1} - \sum_{i=0}^{N-1}\left[i^2 - (N-1)\cdot i\right]\left(C_i + C_i^T\right)\right] =$$
$$= \frac{8}{N^3}\cdot\left[\{i^2\}_{i,j=0}^{N-1} + \{j^2\}_{i,j=0}^{N-1} - 2\{i\cdot j\}_{i,j=0}^{N-1} - \sum_{i=0}^{N-1}\left[i^2 - (N-1)\cdot i\right]\left(C_i + C_i^T\right)\right] =$$
$$= \frac{8}{N^3}\cdot\left[(N-1)\{i\}_{i,j=0}^{N-1} + (N-1)\{j\}_{i,j=0}^{N-1} - 2\{i\cdot j\}_{i,j=0}^{N-1}\right] =$$
$$= \frac{8}{N^3}\cdot\left[\{[(N-1)-2j]\cdot i\}_{i,j=0}^{N-1} + (N-1)\{j\}_{i,j=0}^{N-1}\right] \tag{C.20}$$

The Rank of $\tilde{D}$ is two, since it is a sum of two matrices, each of Rank=1.

The sum of the two nonzero eigenvalues is:

$$\omega_0 + \omega_a = trace\{\tilde{D}\} = \frac{16(N-1)}{N^3}\sum_{i=0}^{N-1} i - \frac{16}{N^3}\sum_{i=0}^{N-1} i^2 = \frac{8(N-1)(N-2)}{3N^2} \tag{C.21}$$

We used the notation $\omega_a$, to point out that the order of the eigenvalues is not yet determined. $\tilde{D}$ was constructed such that the vector $\underline{1}$ is one of its eigenvectors. In order to calculate $\omega_0$, only the sum of the first line ($i = 0$) is needed:

$$\begin{aligned} \omega_0 &= \frac{8(N-1)}{N^3} \cdot \sum_{j=0}^{N-1} j = \frac{4(N-1)^2}{N^2} \\ & \qquad\qquad \Leftarrow (C.21) \\ \omega_a &= -\frac{4(N-1)(N+1)}{3N^2} \end{aligned} \tag{C.22}$$

Note that $\omega_a$ is strictly negative. After sorting the eigenvalues, it will be placed as $\omega_{N-1}$.

The eigenvalues of $Q_N$ are know to be $(1-2q)^m$, $m = 0,1,\ldots,L$, (B.23), where $L$ is the number of bits. The lower and upper bounds are therefore:

Lower bound:

$$D_c \geq \frac{1}{2}\left[\lambda_0 + \lambda_1(1-2q)\right] - S = \frac{2(N-1)(N+1)}{3N^2} \cdot 2q$$

Upper bound:

$$D_c \leq \frac{1}{2}\left[\lambda_0 + \lambda_1(1-2q)^L\right] - S = \frac{2(N-1)(N+1)}{3N^2}\left[1-(1-2q)^L\right] \tag{C.23}$$

as found in (B.24).

◊

# Appendix D                                                    נספח ד'

## האופטימליות של הקוד הבינארי הטבעי עבור קוונטייזר סקלרי אחיד, מקור אחיד ותחת הערוץ הבינארי הסימטרי עם קוד Hamming (7,4) לתיקון שגיאות

## The optimality of the Natural Binary Code Assignment for a Uniform Scalar Quantizer and a Uniform Source under the Binary Symmetric Channel with a (7,4) Hamming error correcting code

In section 6.3.2 we consider the special case of a uniform scalar quantizer and a uniform source under the Binary Symmetric Channel (BSC), with (7,4) Hamming error correcting code [Ta86 Section 13.17], it is shown here that the Natural Binary Code (NBC) offers the minimal Channel Distortion over all possible code assignments.

For an $L$-bit ($N = 2^L$ reconstruction levels) uniform scalar quantizer, we organize the $N$ reconstruction levels in a column vector (B.1):

$$\underline{\rho} = \begin{bmatrix} \rho_0 \\ \rho_1 \\ \vdots \\ \rho_i \\ \vdots \\ \rho_{N-1} \end{bmatrix} = \frac{2}{N} \cdot \begin{bmatrix} -N/2+1/2 \\ 1-N/2+1/2 \\ \vdots \\ i-N/2+1/2 \\ \vdots \\ N/2-1+1/2 \end{bmatrix} \qquad \text{(D.1)}$$

In appendix B we proved that the NBC is the optimal assignment if the vector $\underline{\rho}$ is an eigenvector of the channel matrix $Q$, corresponding to the largest eigenvalue except for 1. In this appendix we show that this is indeed the case, also for $Q_H$ - the channel matrix for the Binary Symmetric Channel, with (7,4) Hamming error correcting code.

The (7,4) Hamming code has a minimum distance of 3 for each pair of codewords, thus is capable of correcting a single bit error.

We denote the codeword of the index $i$ by $c(i)$.

The codewords are:

| Index $i$ | Codeword $c(i)$ | Index $i$ | Codeword $c(i)$ |
|---|---|---|---|
| 0 | 0000000 | 8 | 1000111 |
| 1 | 0001110 | 9 | 1001001 |
| 2 | 0010101 | 10 | 1010010 |
| 3 | 0011011 | 11 | 1011100 |
| 4 | 0100011 | 12 | 1100100 |
| 5 | 0101101 | 13 | 1101010 |
| 6 | 0110110 | 14 | 1110001 |
| 7 | 0111000 | 15 | 1111111 |

Table D.1 - (7,4) Hamming Error Correcting Code

When a codeword $c(i)$ is transmitted, every 7-bit number $k$ has a probability of $q^{H(k,c(i))}(1-q)^{7-H(k,c(i))}$ to appear at the receiver, where $H(k,c(i))$ is the Hamming distance between the codeword and the received number.

We examine a single entry of the channel transition matrix $Q_H$. Assume the VQ Encoder needs to transmit the index $i$. The Hamming codeword $c(i)$ is sent through the BSC. There are 8 possible BSC outputs that will cause the index $j$ to appear at the receiver. These are the Hamming codeword $c(j)$ and its Hamming-1 distance neighbors. Each entry of $Q_H$ is therefore a sum of 8 probabilities:

$$\{Q_H\}_{i,j} = \text{Prob}\{\text{VQ Index } j \text{ received} \mid \text{VQ Index } i \text{ transmitted}\} =$$
$$= \sum_{H(k,c(j))\leq 1} \text{Prob}\{k \text{ received} \mid \text{codeword } c(i) \text{ transmitted}\} \tag{D.2}$$

The channel matrix $Q_H$ has the following form:

$$Q_H = \underbrace{a \cdot I_{16}}_{1} + \underbrace{d \cdot J_{16}}_{2} + \underbrace{\begin{bmatrix} A & B & A & B \\ B & A & B & A \\ A & B & A & B \\ B & A & B & A \end{bmatrix}}_{3} \tag{D.3}$$

where:

$$I_{16} = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix} \; ; \; J_{16} = \begin{bmatrix} 0 & & & 1 \\ & & 1 & \\ & \iddots & & \\ 1 & & & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} c & b & b & c \\ b & c & c & b \\ b & c & c & b \\ c & b & b & c \end{bmatrix} \; ; \; B = \begin{bmatrix} b & c & c & b \\ c & b & b & c \\ c & b & b & c \\ b & c & c & b \end{bmatrix}$$

with:

$$\begin{aligned} a &= q_0 + 7q_1 - 4q_3 - q_4 - 3q_5 \\ b &= 3q_2 + q_3 + 4q_4 \\ c &= 4q_3 + q_4 + 3q_5 \\ d &= 7q_6 + q_7 - 3q_2 - q_3 - 4q_4 \end{aligned} \tag{D.4}$$

and $q_i = q^i(1-q)^{7-i}$ , $i = 0,1,\ldots,7$ is the probability of error in $i$ bits of the codeword.

We now show that the columns of the Walsh-Hadamard matrix as described in Appendix B are the eigenvectors of each the three components of the matrix $Q_H$, denoted 1,2 and 3 in (D.3).

We use a similar recursion methods as in (B.22).

Matrix 1: All columns are eigenvectors of the unity matrix $I$ with a corresponding eigenvalue 1, since $I \cdot \underline{v} = \underline{v}$ for any vector $\underline{v}$.

Matrix 2: For a reverse-diagonal matrix of size 2, the following eigenvectors are verified:

$$\begin{aligned} J_2 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ J_2 \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -1 \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} \end{aligned} \tag{D.5}$$

Now assume the vector $\underline{v}$ is an eigenvector of a reverse-diagonal matrix of size $N$, with a corresponding eigenvalue $\alpha$. We can construct two eigenvectors of a reverse-diagonal matrix of size $2N$, by using:

$$\begin{aligned}&\text{Duplication:}\\ &J_{2N}\cdot\begin{bmatrix}\underline{v}\\ \underline{v}\end{bmatrix}=\begin{bmatrix}0 & J_N\\ J_N & 0\end{bmatrix}\cdot\begin{bmatrix}\underline{v}\\ \underline{v}\end{bmatrix}=\alpha\cdot\begin{bmatrix}\underline{v}\\ \underline{v}\end{bmatrix}\\ &\text{Negative Duplication:}\\ &J_{2N}\cdot\begin{bmatrix}\underline{v}\\ -\underline{v}\end{bmatrix}=\begin{bmatrix}0 & J_N\\ J_N & 0\end{bmatrix}\cdot\begin{bmatrix}\underline{v}\\ -\underline{v}\end{bmatrix}=-\alpha\cdot\begin{bmatrix}\underline{v}\\ -\underline{v}\end{bmatrix}\end{aligned} \tag{D.6}$$

Therefore, the matrix $J_{16}$ has two eigenvalues: $\pm 1$, each with multiplicity 8. A Walsh-Hadamard eigenvector has a corresponding positive eigenvalues if it is constructed using an even number of negative duplications.

Matrix 3: The third matrix is constructed of two types of building blocks, both of size 4 and sharing the same eigenvectors. Their corresponding eigenvalue-eigenvectors are:

$$\begin{aligned}&A\cdot\begin{bmatrix}1\\1\\1\\1\end{bmatrix}=2(b+c)\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\;;\quad A\cdot\begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}=0\cdot\begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}\\ &A\cdot\begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}=0\cdot\begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}\;;\quad A\cdot\begin{bmatrix}1\\-1\\-1\\1\end{bmatrix}=2(c-b)\cdot\begin{bmatrix}1\\-1\\-1\\1\end{bmatrix}\end{aligned} \tag{D.7a}$$

$$\begin{aligned}&B\cdot\begin{bmatrix}1\\1\\1\\1\end{bmatrix}=2(b+c)\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\;;\quad B\cdot\begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}=0\cdot\begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}\\ &B\cdot\begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}=0\cdot\begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}\;;\quad B\cdot\begin{bmatrix}1\\-1\\-1\\1\end{bmatrix}=2(b-c)\cdot\begin{bmatrix}1\\-1\\-1\\1\end{bmatrix}\end{aligned} \tag{D.7b}$$

Now assume that the vector $\underline{v}$ is an eigenvector of both A and B, with a corresponding eigenvalues $\alpha$ and $\beta$ respectively. We can construct four eigenvectors for matrix 3 (size 16).

$$\begin{bmatrix} A & B & A & B \\ B & A & B & A \\ A & B & A & B \\ B & A & B & A \end{bmatrix} \cdot \begin{bmatrix} \underline{v} \\ \underline{v} \\ \underline{v} \\ \underline{v} \end{bmatrix} = 2(\alpha+\beta) \begin{bmatrix} \underline{v} \\ \underline{v} \\ \underline{v} \\ \underline{v} \end{bmatrix} \; ; \; \begin{bmatrix} A & B & A & B \\ B & A & B & A \\ A & B & A & B \\ B & A & B & A \end{bmatrix} \cdot \begin{bmatrix} \underline{v} \\ -\underline{v} \\ \underline{v} \\ -\underline{v} \end{bmatrix} = 2(\alpha-\beta) \begin{bmatrix} \underline{v} \\ -\underline{v} \\ \underline{v} \\ -\underline{v} \end{bmatrix}$$

$$\begin{bmatrix} A & B & A & B \\ B & A & B & A \\ A & B & A & B \\ B & A & B & A \end{bmatrix} \cdot \begin{bmatrix} \underline{v} \\ \underline{v} \\ -\underline{v} \\ -\underline{v} \end{bmatrix} = 0 \cdot \begin{bmatrix} \underline{v} \\ \underline{v} \\ -\underline{v} \\ -\underline{v} \end{bmatrix} \; ; \; \begin{bmatrix} A & B & A & B \\ B & A & B & A \\ A & B & A & B \\ B & A & B & A \end{bmatrix} \cdot \begin{bmatrix} \underline{v} \\ -\underline{v} \\ -\underline{v} \\ \underline{v} \end{bmatrix} = 0 \cdot \begin{bmatrix} \underline{v} \\ -\underline{v} \\ -\underline{v} \\ \underline{v} \end{bmatrix} \qquad \text{(D.8)}$$

We apply (D.7) and (D.8) to construct the 16 eigenvectors of Matrix 3 of (D.3). Matrix 3 has the Walsh-Hadamard eigenvectors with the following corresponding eigenvalues:

a. $8(c+b) \Leftrightarrow [1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1]^T$

b. $8(c-b) \Leftrightarrow [1 \;\; -1 \;\; -1 \;\; 1 \;\; -1 \;\; 1 \;\; 1 \;\; -1 \;\; 1 \;\; -1 \;\; -1 \;\; 1 \;\; -1 \;\; 1 \;\; 1 \;\; -1]^T$ (D.9)

The eigenvalue zero has a multiplicity of 14.

Using all the above information we may list the eigenvalues and corresponding eigenvectors of the channel matrix $Q_{II}$:

- There is one eigenvector (D.9a) constructed with no negative-duplication, and has eigenvalue 1.
- There are four eigenvectors constructed with one negative-duplication, all have eigenvalue $(a-d)$.
- There are six eigenvectors constructed with two negative-duplication, all have eigenvalue $(a+d)$. Note that $d$ is negative, therefore this eigenvalue is smaller than $(a-d)$.

- There are four eigenvectors constructed with three negative-duplication, three have eigenvalue $(a-d)$. One eigenvector, (D.9b) has a corresponding eigenvalue $(a-d)+8(c-b)$. Note that $b>c$, therefore this eigenvalue is smaller than $(a-d)$.

  There is one eigenvector constructed with four negative-duplication, and has eigenvalue $(a+d)$.

In conclusion, the channel matrix $Q_H$ has four distinct eigenvalues (in decreasing order): 1, $(a-d)$, $(a+d)$, $(a-d)+8(c-b)$ . The largest, except for 1, is $(a-d)$ and it corresponds to Walsh-Hadamard eigenvectors constructed with a single negative-duplication. In Appendix B, (B.30), we show that the representing-levels vector $\underline{\rho}$ is a linear combination of these eigenvectors.
This implies that $\underline{\rho}$ is an eigenvector of $Q_H$ corresponding to the eigenvalue $(a-d)$.
This is the property that was needed to be verified. Therefore, the Natural-Binary-Code is also the optimal assignment for the Uniform Scalar Quantizer, Uniform Source and a Binary Symmetric Channel using the (7,4) Hamming Code.

# Appendix E                                                    נספח ה'

## אלגוריתם לשיבוץ תת-אופטימלי של אינדקסים בקוונטייזר וקטורי עבור ערוצים סימטריים חסרי-זיכרון

## Algorithm for Suboptimal Index Assignment in Vector Quantizers for Symmetric Memoryless Channels

In this appendix we present an algorithm for suboptimal index assignment for Vector Quantizers, operating under Symmetric Memoryless Channels. The algorithm is also described in Chapter 7, and numerical results are presented in chapter 6. We assume knowledge of the Channel transition matrix $Q_N$, the source probabilities matrix $P$, and the VQ codevectors distance matrix $D$.

The optimization problem is, from (2.15):

$$\min_{\pi} D_c(\pi) = trace\{P\pi Q_N \pi^T D\} \tag{E.1}$$

where $\pi$ is an unknown permutation matrix.

In Appendix C, and in Section 5.3.2, we have shown that solving (E.1) is equivalent to solving the following optimization problem:

$$\min_{\pi} \quad trace\{Q_N \pi^T \tilde{D} \pi\} \tag{E.2}$$

where $Q_N$ is the channel probability matrix and $\tilde{D}$ is a symmetric matrix, having $\underline{1}$ as an eigenvector. $\tilde{D}$ was constructed from both $P$ and $D$, using an algorithm described as phase 2 in Appendix C.

The symmetric channel probability matrix $Q_N$ has a unitary diagonalization:

$$Q_N = V \cdot \Lambda \cdot V^T \qquad V \cdot V^T = I \tag{E.3}$$

We arrange the eigenvalues in $\Lambda$ to be in decreasing order, with the first column of $V$ being $\frac{1}{\sqrt{N}} \cdot \underline{1}$, and the first eigenvalue is $\lambda_0 = 1$.

Substiuting (E.3) into (E.2):

$$\min_{\pi} \ trace\{V\cdot\Lambda\cdot V^T\pi^T\tilde{D}\pi\} = \min_{\pi} \ trace\{\pi V\cdot\Lambda\cdot V^T\pi^T\tilde{D}\} \tag{E.4}$$

The symmetric matrix $\tilde{D}$ has also a unitary diagonalization:

$$\tilde{D} = Z\cdot\Omega\cdot Z^T \tag{E.5}$$

with the eigenvalues $\omega_i$ and corresponding eigenvectors $\underline{z}_i$. Moreover, $\tilde{D}$ was constructed to have the eigenvector $\underline{z}_0 = \frac{1}{\sqrt{N}}\underline{1}$ with a corresponding eigenvalue $\omega_0$. The eigenvalue $\omega_0$ is the Perron-Frobenius eigenvalue of $\tilde{D}$. Therefore it is positive and largest in absolute value. All eigenvalues are arranged in decreasing order.

The optimization problem is now:

$$\min_{\pi} \ trace\{\Lambda V^T\pi^T Z\Omega Z^T\pi V\} =$$

$$\Leftarrow trace\{AB\} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} A_{ij}^T B_{ij}$$

$$\min_{\pi} \ \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\lambda_j\omega_i\left(Z^T\pi V\right)_{ij}^2 \tag{E.6}$$

The matrix $\pi V$ is unitary for every permutation-matrix $\pi$, since $(\pi V)(\pi V)^T = \pi VV^T\pi^T = \pi\pi^T = I$. Moreover, for the first eigenvalue of $Q_N$, $\lambda_0 = 1$, permutation does not affect its corresponding eigenvector: $\pi\underline{v}_0 = \frac{1}{\sqrt{N}}\pi\underline{1} = \frac{1}{\sqrt{N}}\underline{1}$.

Define a matrix $\Psi$ as $\Psi = Z^T\pi V$. This matrix is also unitary for every permutation-matrix $\pi$, since $\Psi\Psi^T = \left(Z^T\pi V\right)\left(Z^T\pi V\right)^T = Z^T\pi VV^T\pi^T Z = Z^T\pi\pi^T Z = Z^T Z = I$.

Since the first column of both $V$ and $Z$ is $\underline{v}_0 = \underline{z}_0 = \frac{1}{\sqrt{N}}\underline{1}$, and the remaining columns are orthogonal to the vector $\underline{1}$, the structure of $\Psi = Z^T \pi V$ is:

$$\Psi = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & ? & \\ 0 & & & \end{bmatrix} \tag{E.7}$$

where the question mark represents unknown entries.

We have seen in Appendix C, (C.12), that if $\pi V$ is replaced by a general unitary matrix, $U$, the minimal target function in (E.6) is achieved for:

$$\Psi^* = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & & & 1 \\ \vdots & & & 1 & \\ 0 & & \iddots & & \\ 0 & 1 & & & 0 \end{bmatrix} \tag{E.8}$$

The matrix $\Psi^*$ is not an outcome of a valid permutation matrix $\pi$ in $\Psi = Z^T \pi V$. Therefore, we state an approximation by a general unitary matrix $W$. Replacing $\pi$ with a general unitary matrix $W$, the minimal value for the target function is achieved for $W = Z\Psi^* V^T$. All entries of $W$ are smaller in absolute value than 1: $|W_{ij}| \le 1, \quad i, j = 0, 1, \ldots, N-1$.

Next, we search for the permutation matrix "closest" to the matrix $W$, by solving a standard Assignment problem:

$$\begin{aligned} \min_{\pi} \quad & \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(1 - w_{ij}^2\right)\pi_{ij} \\ s.t. \quad & \sum_{j=0}^{N-1} \pi_{ij} = 1 \quad i = 0, 1, \ldots, N-1 \\ & \sum_{i=0}^{N-1} \pi_{ij} = 1 \quad j = 0, 1, \ldots, N-1 \\ & \pi_{ij} \ge 0 \end{aligned} \tag{E.9}$$

The Weighting function $\left(1-w_{ij}^{2}\right)$ is one of several possible weighting function and was found to perform well for the numerical examples we checked. Solving the Assignment problem (E.9), can be done by the Hungarian Method [Ta87 Section 6.3] known from Operations Research.

The outcome of the Assignment problem is a permutation matrix that can be used in the target function (E.1)

**Remark 1:**

For the Binary-Symmetric-Channel and small values of the Bit Error Rate $q$, it is possible to derive an asymptotically suboptimal index assignment by solving an optimization problem similar to (E.4). The result is an index assignment for an interval of channel Bit Error Rate: $0 \le q \le q_0$. The eigenvalues of the Channel matrix are $(1-2q)^l$, $l = 0,1,\dots,L$, (B.23). In order to obtain the asymptotic behavior of these eigenvalues, we note that: $\lim\limits_{q\to 0}\dfrac{1}{-2q}\left((1-2q)^l - 1\right) = l$.

We therefore replace the diagonal eigenvalues matrix $\Lambda$ with an asymptotic diagonal matrix $\hat{\Lambda} = \lim\limits_{q\to 0}\dfrac{1}{-2q}(\Lambda - I)$.

The optimization solution is not affected by an addition of a constant (independent of the permutation matrix $\pi$) to the target function. Moreover, multiplying the target function by a negative constant will change the minimization problem to a maximization one.

The optimization problem, for small Bit Error Rates $q \to 0$, is changed as follows:

$$\min_{\pi} \quad trace\left\{\pi V \cdot \Lambda \cdot V^T \pi^T \tilde{D}\right\}$$

$$\therefore \max_{\pi} \quad \left(-\frac{1}{2q} trace\left\{\pi V \cdot \Lambda \cdot V^T \pi^T \tilde{D}\right\}\right)$$

$$\therefore \max_{\pi} \quad \left(-\frac{1}{2q} trace\left\{\pi V \cdot \Lambda \cdot V^T \pi^T \tilde{D}\right\} - trace\left\{\tilde{D}\right\}\right)$$

$$\therefore \max_{\pi} \quad \left(-\frac{1}{2q} trace\left\{\pi V \cdot (\Lambda - I) \cdot V^T \pi^T \tilde{D}\right\}\right)$$

$$\therefore \max_{\pi} \quad \left(trace\left\{\pi V \cdot \left[\lim_{q \to 0}\left(-\frac{1}{2q}(\Lambda - I)\right)\right] \cdot V^T \pi^T \tilde{D}\right\}\right)$$

$$\therefore \max_{\pi} \quad trace\left\{\pi V \cdot \hat{\Lambda} \cdot V^T \pi^T \tilde{D}\right\} \tag{E.10}$$

Note that for small values of $q$, the maximization in (E.10) is independent of the actual value of $q$.

Therefore, in order to obtain an asymptotically suboptimal (robust) index assignment, use the optimization method described starting in (E.4) to solve the problem: $\max_{\pi} \quad trace\left\{\pi V \cdot \hat{\Lambda} \cdot V^T \pi^T \tilde{D}\right\}$.

**Remark 2 (complexity):**

The complexity of the described algorithm is $O(N^3)$, since its phases are as follows:

1. Calculation of Eigenvalues and Eigenvectors for two Hermitian Matrices [Go89 Section 8.2]: $O(N^3)$
2. Matrix multiplication: $O(N^2)$
3. Solving a standard Assignment Problem [Ta87 Section 6.3]: $O(N^3)$

# Appendix F

# נספח ו'

## תכנות ליניארי - הגדרות והמשפט היסודי

בעבודה זו, בפרקים 3 ו-7, נעשה שימוש בטכניקות של תכנות ליניארי (LP - Linear Programming). ניתן למצוא חומר מקיף בנושא בספרי הלימוד [Ta87], [Lu84]. בנספח זה נביא את ההגדרה של בעיית התכנות הליניארי, מונחים בסיסיים ואת המשפט היסודי של התכנות הליניארי. פתרון בעיית התכנות הליניארי נעשה ע"י אלגוריתם ה-Simplex ונגזרותיו.

בעיית תכנות ליניארי היא בעיית אופטימיזציה וקטורית, בה גם פונקצית המטרה וגם האילוצים הם ליניאריים במשתנים. הבעיה יכולה להיות בעיית מינימיזציה או מקסימיזציה, עם אילוצי שוויון ואי-שוויון:

$$\begin{aligned} \min_{\underline{x}} / \max_{\underline{x}} \quad & \underline{c}^T \cdot \underline{x} \\ \text{s.t.} \quad & A_1 \cdot \underline{x} = \underline{b}_1 \\ & A_2 \cdot \underline{x} \geq \underline{b}_2 \\ & A_3 \cdot \underline{x} \leq \underline{b}_3 \end{aligned} \tag{F.1}$$

כאשר המטריצות $A_1, A_2, A_3$ והוקטורים $\underline{c}, \underline{b}_1, \underline{b}_2, \underline{b}_3$ הנם בממדים מתאימים.

ניתן להפוך באמצעות אלגברה פשוטה כל בעיית תכנות ליניארי לבעיית **תכנות ליניארי בצורה סטנדרטית** :

$$\begin{aligned} \min_{\underline{x}} \quad & \underline{c}^T \cdot \underline{x} \\ \text{s.t.} \quad & A \cdot \underline{x} = \underline{b} \\ & \underline{x} \geq \underline{0} \end{aligned} \tag{F.2}$$

כאשר $\underline{0}$ הוא וקטור אפסים. מימדי המטריצות והוקטורים הם:

$$\begin{aligned} & \underline{c}^{n \cdot 1}, \quad \underline{x}^{n \cdot 1}, \quad \underline{b}^{m \cdot 1}, \quad A^{m \cdot n} \\ & \text{and} \quad m < n \end{aligned} \tag{F.3}$$

$A$ היא מטריצה עם מימד מלא ($Rank(A) = m$).

מכיוון שניתן להפוך כל בעיית תכנות ליניארי לצורה הסטנדרטית, מספיק להציע אלגוריתם לפתרון הצורה הזו.

נסתכל על מערכת המשוואות:

$$A \cdot \underline{x} = \underline{b} \tag{F.4}$$

למערכת המשוואות יש $n$ נעלמים ב-$m$ משוואות ($m < n$), ולכן יש למערכת אינסוף פתרונות אפשריים.

**<u>הגדרה</u>**: פתרון קביל (Feasible Solution) - פתרון המקיים את האילוצים : $A\underline{x}=\underline{b}$ ו-$\underline{x}\geq 0$.

**<u>משפט</u>**: קבוצת הפתרונות הקבילים הנה קבוצה קמורה (Convex Set) [Lu84 סעיף 2.5].

**<u>הגדרה</u>**: פתרון בסיסי (Basic Solution) - פתרון למערכת המשוואות (F.4) שבו לפחות $n-m$ משתנים מן הוקטור $\underline{x}$ שווים לאפס.

ניתן למצוא את $\binom{n}{m}$ הפתרונות הבסיסיים ע״י בחירת $n-m$ משתנים מן הוקטור $\underline{x}$ והשוואתם לאפס. שאר $m$ המשתנים נקבעים ע״י פתרון מערכת המשוואות ($m$ משוואות) הנותרת. כאשר אין למערכת המשוואות פתרון יחיד, מתקבל **פתרון בסיסי מנוון**. הטיפול בפתרונות מנוונים הוא מחוץ למטרת נספח זה.

**<u>הגדרה</u>**: פתרון בסיסי קביל (Feasible Basic Solution) - פתרון בסיסי המקיים גם את אילוץ ה-״אי-שליליות״ : $\underline{x}\geq 0$.

לבעיית התכנות הליניארי ייתכנו המקרים הבאים :

1. **אין פתרון קביל**, כלומר מערכת האילוצים כוללת סתירות.
2. **יש למערכת פתרונות קבילים אך אין לה פתרונות אופטימליים**. לדוגמה עבור המערכת :

$$\begin{aligned} \min_{\underline{x}} \quad & x_1 - x_2 + x_3 \\ \text{s.t.} \quad & x_1 + 2x_3 = 17 \\ & x_1, x_2, x_3 \geq 0 \end{aligned} \tag{F.5}$$

אין אילוץ שוויון על $x_2$ והגדלתו תקטין את פונקצית המטרה למינוס אינסוף.

3. **יש למערכת פתרון קביל ואופטימלי.**

**<u>המשפט היסודי של התכנות הליניארי</u>**:

עבור בעיית תכנות ליניארי במבנה סטנדרטי (F.2) :

1. אם קיים פתרון קביל אזי קיים גם פתרון בסיסי קביל.
2. אם קיים פתרון קביל אופטימלי אזי קיים גם פתרון בסיסי קביל אופטימלי.

**<u>הערה</u>**: ייתכן מקרה בו קיים פתרון אופטימלי לא בסיסי, אולם המשפט מבטיח כי קיים פתרון בסיסי עם ערך פונקצית מטרה זהה.

המשפט היסודי מבטיח כי מספיק לחפש את הפתרון האופטימלי בפתרונות הבסיסיים ש<u>מספרם סופי</u>. אלגוריתם ה-Simplex מבצע בכל איטרציה ״קפיצה״ בין פתרון בסיסי קביל לפתרון בסיסי קביל אחר עם ערך פונקציה מטרה נמוך יותר. ובכך מגיע תוך מספר צעדים סופי לאופטימום.

# VECTOR AND SCALAR QUANTIZATION UNDER CHANNEL-ERRORS

by
Gal Ben-David

# VECTOR AND SCALAR QUANTIZATION UNDER CHANNEL-ERRORS

Research Thesis

Submitted in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Science

by
Gal Ben-David

Submitted to the Senate of the Technion - Israel Institute of Technology

Sivan 5755 Haifa June 1995

**The work described herein was supervised by Prof. David Malah in the Faculty of Electrical Engineering**

My sincere expression of gratitude to my supervisor, Prof. David Malah, for proposing the research subject and for his devoted guidance.

This paper concludes a wonderful period in which I earned my education in the Signal Processing Laboratory.
My deep thanks to Yoram Or-Chen for my first project and for his support.
I would like to thank Ziva, Nimrod, Uri Avi and Tamara for the great time we spent together.

I would like to thank to the Wolf foundation for the generous support by the Dr. Moshe Gilboa Memorial scholarship

To Renata and Dan-Ilan

# Table of Contents

## Table of Contents (continued)

# Table of Contents (continued)

# Table of Contents (continued)

# Abstract


Vector Quantization (VQ) systems (or Scalar Quantization as a special case) are widely implemented in Source-Coding Applications. VQ can be found in many modern speech and image coding systems. A Vector Quantizer divides the set of all possible input signals (vectors) into a finite set of Partition Cells, where each cell is represented by a Representation Vector. VQ is a method for Lossy Compression. Due to quantization, part of information existing in the original signal is lost, and the reconstructed signal is not identical to the original. The difference between the original signal and its reconstruction is known as the Quantization Distortion.

A VQ based communication system consists of two parts: Encoder and Decoder. The Encoder (prior to the digital channel) examines the input vector and decides to which partition cell it belongs. Both the Encoder and Decoder keep an identical table of Representation Vectors. Each vector is assigned with a Channel Index. This index is transmitted through the channel, and the Decoder pulls out a corresponding Representation Vector.

VQ design is based on the knowledge of the statistics of the signal source. In practice, the design is based on a Training Sequence collected at the source. "Classical" analysis and design methods of VQ systems assume that no error occurs in the digital channel. For practical channels, a channel Error-Correcting-Code is used, which in many situations cause the Channel-Errors effects to be negligible.

In many other cases the effect of Channel Errors is significant, and better performance can be obtained by using Vector Quantizers that take channel errors into consideration.

One of the goals of this work is to examine the effect of channel errors on the performance of given VQ systems. A channel error causes an incorrect index to appear at the decoder. The decoder then reconstructs an incorrect vector. The distance between the reconstructed vector and the vector that should have been reconstructed, is known as the Channel Distortion. The distance between the (erroneous) reconstructed vector and the original input vector is the Total (Overall) distortion. In this work we assume a memoryless channel, while some of the results are being

applicable only to the special, yet useful, case of the Binary-Symmetric-Channel (BSC).

In addition to finding bounds on the performance of given VQ systems, two factors affecting Channel distortion are examined. The first is the constellation of the Vector (and Scalar) Quantizer in terms of Partition Cells and Representation Vectors. The other element is the assignment of channel indices to the Reconstruction Vectors.

In the first part of the this work upper and lower bounds on the Channel-Distortion over all possible input sources are obtained. Several numerical examples are shown for demonstration. For the special case of a Uniform Scalar Quantizer (USQ), Natural Binary Code (NBC) and a BSC, the upper bound coincides with the lower bound. Moreover, for the same case, the Channel Distortion is practically independent of the number of bits of the quantizer. In other cases the gap between the upper and lower bounds could be large and their knowledge does not give enough information to estimate the system performance. For these cases, tighter bounds are presented for a reduced set of possible sources, characterized by source moments, and especially power constraints. Asymptotical bounds for the BSC for small channel bit error rate ($q \to 0$) are also presented.

For USQ of 4-bits and more, we also show that the channel distortion due to the NBC is better than the average distortion over all possible index assignments for all possible source.

In previous works found in the literature, necessary conditions for optimal VQ design in the presence of Channel-Errors are given. These conditions give rise to a complicated design procedure, producing a library of different quantizers, each for a different error rate of the channel. Storing such a library, in both the Encoder and Decoder, results in costly memory requirements. In this work we present a simple suboptimal approach. We suggest to linearly scale the Partition Cells and Reconstruction Vectors as a function of the channel error rate. The design procedure is much simpler, most of it coincides with the classical (no channel errors) design. The memory requirements of the proposed system is much less demanding. Numerical examples are shown, demonstrating that the performance loss compared with the optimal approach is relatively small.

An important factor influencing Channel Distortion is the assignment of indices to the reconstruction vectors. Generally, the goal is to make channel errors do as less damage as possible. There is a huge number of possible index assignments. For a VQ with $N$ representation vectors there are $N!$ possible assignments, causing an exhaustive search over all possible assignment to be practically impossible.

For the special case of the USQ, Uniform Source, and the BSC, the NBC assignment is shown to be optimal.

Lower and Upper bounds on the Channel Distortion over all possible assignments for any Memoryless channel are then presented.

We conclude this work by developing a suboptimal algorithm for index assignment for a VQ operating under Symmetric Memoryless Channels. The complexity of the algorithm is shown to be of $O(N^3)$, and it performs better than the most extensively used algorithm - Index Switching, whose complexity is $O(N^4)$.

In numerical examples, the bounds over all possible index assignments are compared with 10,000 random assignments, and for two cases also with the proposed suboptimal assignment. The limited random search is found to be insufficient due to the vast number of possible assignments (for a merely 4-bit quantizer we examined $10^4$ assignment out of the possible $2 \cdot 10^{13}$). For the special case of USQ and a uniform source, the lower bound coincides with the optimal assignment.

For a typical case examined by simulation, the Channel-Distortion due to the assignment given by the proposed algorithm is about 8dB lower than the average distortion over all possible index assignments. For the same case, the distortion due to the index switching algorithm is about 1dB higher than the proposed algorithm.