دانشکده علوم ریاضی

پایان‌نامه کارشناسی ارشد ریاضی کاربردی

کدگذاری سیستم‌های بیولوژیکی و شبکه‌های تنظیم ژن

نگارنده: الهام بهرامی‌پور

استادان راهنما

دکتر عبدالله آل‌هوز
دکتر میثم علیشاهی

استاد مشاور
دکتر ابراهیم هاشمی

مهرماه ۱۳۹۹

این پایان‌نامه را تقدیم می‌کنم به:
همسر مهربانم،
که وجودش تعریف کامل انسانیت و نشانه لطف الهی است،
پسرم محمدکیان،
و پدر و مادر بزرگوارم...

سپاس‌گزاری

از آنجایی که تجلیل از معلم، سپاس از انسانی است که هدف و غایت آفرینش را تأمین می‌کند. برحسب وظیفه از زحمات بی‌دریغ اساتید ارجمندم جناب آقای دکتر آل‌هوز که زحمت راهنمایی این پایان‌نامه را به عنوان استاد راهنمای اول کشیدند و جناب آقای دکتر علیشاهی به عنوان استاد راهنمای دوم و جناب آقای دکتر هاشمی به عنوان مشاور، کمال قدردانی و تشکر را دارم.

تعهد نامه

اینجانب **الهام بهرامی‌پور** دانشجوی کارشناسی ارشد رشته **ریاضی کاربردی** دانشکده **علوم ریاضی** دانشگاه شاهرود، نویسنده پایان‌نامه با عنوان **کدگذاری سیستم‌های بیولوژیکی و شبکه‌های تنظیم ژن**، تحت راهنمایی دکتر عبدالله آل‌هوز و دکتر میثم علیشاهی متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های دیگر پژوهش‌گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان‌نامه، تا کنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ‌جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه صنعتی شاهرود تعلق دارد، و مقالات مستخرج با نام " دانشگاه صنعتی شاهرود " یا " Shahrood University of Technology " به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به‌دست آوردن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند، در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده شده است)، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

الهام بهرامی‌پور
مهرماه ۱۳۹۹

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می‌باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان‌نامه بدون ذکر منبع مجاز نمی‌باشد.

چکیده

امروزه استفاده از علوم ریاضی در زیست‌شناسی، از جایگاه ویژه‌ای برخوردار بوده و محک علمی بودن پژوهش‌ها قلمداد می‌شود. ارائه تئوری‌های ریاضی، آمار و احتمال، مطالعه نظری بیماری‌ها با توجه به اختلالات ایجاد شده در مسیر متابولیک یا تغییرات ژنتیکی و ساخت الگوریتم محاسباتی و مدل‌سازی برای پاسخ مسائل مطرح علوم زیست‌شناسی، از شاخص‌ترین کاربردهای ریاضیات در علوم زیستی است. ما در این پایان‌نامه قصد داریم درمورد کاربرد ریاضی در زیست‌شناسی علی‌الخصوص تئوری کدگذاری در زیست‌شناسی سلولی و مولکولی بحث کنیم. ابتدا کدهای دوری اریب، یعنی کدهای خطی روی میدان‌های متناهی که به وسیله ایده‌آل‌های چپ حلقه خارج‌قسمتی چندجمله‌ای‌های اریب ناجابه‌جایی به دست می‌آیند را مورد مطالعه قرار می‌دهیم.

سپس در مورد ساختار کدهای برگشت‌پذیر DNA در $F_{۴۲_s}$ بحث می‌کنیم. همچنین بررسی می‌کنیم که چطور وجود و خواص کدهای دوری اریب به ساختار کدهای DNA برگشت‌پذیر ارتباط پیدا می‌کند.

در نهایت نشان می‌دهیم که دوگان کدهای دوری اریب نیز کدهای DNA برگشت‌پذیرند و چندجمله‌ای خود متقابل اریب نیز یک چندجمله‌ای θ ـ پالیندرومیک است.

کلمات کلیدی: کد ژنتیک، نوکلئوتید، میدان متناهی، کدگذاری کانال، قاعده کدگشایی، کدهای دوری، کد دوری اریب، کدهای DNA برگشت‌پذیر، چندجمله‌ای θ ـ پالیندرومیک، چندجمله‌ای خودمتقابل اریب.

پیشگفتار

در نگاه اول، ممکن است علم ریاضی (نظریه کدگذاری) و علم زیست‌شناسی (سلولی و مولکولی) دو مبحث کاملاً جدا از هم به نظر آیند، که اولی در علوم ریاضیات و دومی در گونه‌های زنده بحث می‌کند، و یا در بهترین حالت ممکن است تصور شود رابطه‌ی این دو یک‌طرفه می‌باشد. نظریه کدگذاری زیستی یک ارتباط عمیق بین ریاضیات و زیست شناسی ایجاد کرده است. علت اصلی پررنگ شدن ریاضیات زیستی در شاخه‌های مختلف، پیشرفت‌های اخیر در علم زیست شناسی است. در گذشته دانشمندان نظم خاصی برای حیات قائل نبودند؛ اما پیشرفت‌های اخیر نظم و مشابهت‌های زیادی را در اجزای حیات کشف کرد. کشف DNA و بررسی ساختار آن نشان داد که عملکرد سلول‌ها، بسیار دقیق و ریاضی وارتر از آن چیزی است که پیش از این به نظر می‌آمد. این پیشرفت ها سنگ بنای ورود جدی‌تر ریاضی دانان به حوزه زیست شناسی بود. اولین پیشگامان این حیطه تحقیقاتی شامل رونالد فیشر[۱] ، ویتو ولترا[۲] و نیکولاس راشورسکی[۳] می‌باشند.

این پایان‌نامه شامل سه فصل است. در فصل اول تعاریف و مفاهیم اولیه مورد نیاز برای ادامه کار بیان می‌شوند. در فصل دوم پس از بیان مفاهیم اولیه به بیان نمایش ریاضی نوکلئوتیدها و مدل‌سازی کانال‌های تکراری می‌پردازیم. در فصل سوم پس از بیان مفاهیم اولیه و اثبات چند لم و قضیه ساختار کدهای دوری اریب در نظریه کدگذاری و DNA در زیست شناسی سلولی و مولکولی را بررسی می‌کنیم.

[1] Ronald Fisher
[2] Vito Voltera
[3] Nicholas Rashorsky

فهرست مطالب

پیشگفتار آ

فهرست تصاویر ه

فهرست جداول ز

۱ تعاریف و مقدمات ۱

۱.۱ مقدمه ۱

۲.۱ تعاریف و مفاهیم لازم از جبر ۲

۳.۱ تعاریف و مفاهیم لازم از نظریه کدگذاری ۱۱

۱.۳.۱ قواعد کدگشایی ۱۳

۲.۳.۱ قاعده کدگشایی ماکزیمم احتمال (MLD) ۱۳

۳.۳.۱ قاعده کدگشایی مینیمم فاصله (MDD) ۱۴

۴.۳.۱ کران‌هایی در نظریه کدگذاری ۱۵

۴.۱ تعاریف و مفاهیم لازم از زیست شناسی سلولی و مولکولی ۱۷

۲ کدگذاری توالی DNA ۲۳

۱.۲ مقدمه ۲۳

۲.۲ نمایش نوکلئوتیدها ۲۴

۳.۲ تعیین کد خطی زمینه ۲۶

۴.۲ مدل‌سازی کانال‌های تکراری ۲۷

۵.۲ الگوریتم رده‌بندی رشته های DNA ۳۰
۶.۲ کشف تکرار و نواحی زائد DNA ۳۲
۷.۲ تست LD ۳۲
۸.۲ شبکه‌های تنظیم ژن ۳۶
۹.۲ کدگذاری و کدگشایی عصبی ۳۸

۳ کدهای DNA برگشت‌پذیر ۴۱
۱.۳ مقدمه ۴۱
۲.۳ دوگان θـ کدهای دوری ۴۸
۳.۳ کدهای DNA برگشت‌پذیر ۵۲
۴.۳ کدهای DNA با طول زوج ۵۲
۵.۳ کدهای DNA با طول فرد ۶۱

مراجع ۶۳

واژه‌نامه فارسی به انگلیسی ۶۹

واژه‌نامه انگلیسی به فارسی ۷۲

فهرست تصاویر

۱.۱ رونویسی و ترجمه در DNA ۱۹

۱.۲ مدل‌سازی کدگذاری و کدگشایی DNA ۲۸

۲.۲ جهش درون قالبی نوکلئوتیدها ۲۹

۳.۲ تست SP بر $(۸,۵)$ کد بلوکی خطی در $GF(۴)$ ۳۱

۴.۲ تست SP بر روی باکتری $E.coli$ ۳۱

۵.۲ تست LD یست کروموزوم ۳۴

۶.۲ تست LD ژنوم انسان ۳۴

۷.۲ الگوریتم LD برای $N = ۴۸$ ۳۵

۸.۲ گراف ژنتیکی ۳۷

فهرست جداول

۱.۲ محاسبات در میدان $GF(۴)$ ۲۶

۱.۳ کدگذاری نوکلئوتیدها ۴۷

فصل ۱

تعاریف و مقدمات

۱.۱ مقدمه

نظریه کدگذاری[۲] یک حوزهٔ مطالعاتی مرتبط به انتقال داده‌ها از طریق کانال پارازیت‌دار و محافظت از آن‌ها در مقابل تحریف‌ها می‌باشد. نظریهٔ کدگذاری مرهون مقاله معروف کلاود شانون[۳] با موضوع ارسال قابل قبول اطلاعات روی کانال‌های پارازیت‌دار است که درسال ۱۹۴۸ مطرح شد. شانون ثابت کرد اگر نرخ ارسال اطلاعات کمتر از ظرفیت کانال باشد، مخابرات مطمئن قابل حصول خواهد بود، به‌شرطی که بتوان از کدگذاری و کدگشایی مناسب استفاده کرد. ابتدا لازم است با توجه به مباحثی که در این پایان‌نامه بررسی می‌کنیم، توضیحی کوتاه راجع به کاربرد نظریه کدگذاری در مولکول DNA ارایه دهیم. فریدریش گاوس[۴] می‌گوید ریاضیات ملکه علوم است. امروزه می‌دانیم علوم بسیاری برای پیشرفت خود به ریاضیات وابسته است. سابقه استفاده از ریاضیات در زیست‌شناسی به قرن دوازدهم برمی‌گردد، زمانی که فیبوناچی[۵] از دنباله معروفش برای توصیف رشد خرگوش‌ها استفاده کرد. علت پررنگ‌تر شدن کاربرد ریاضیات در زیست‌شناسی، پیشرفت‌های اخیر در علم زیست‌شناسی، به ویژه کشف DNA و بررسی ساختار مارپیچ و منظم DNA توسط واتسون و کریک[۶]، نظم و مشابهت‌های زیادی را در اجزای حیات به وجود آورد. از آنجایی که مولکول DNA نقش مهمی در انتقال اطلاعات در سلول دارد و با تکثیر متوالی دچار خطا می‌شود؛ با الهام گرفتن از تصحیح خطا در

---

[2] Coding theory
[3] Claude Shannon
[4] Friedrich Gauss
[5] Fibonacci
[6] Watson-Crick

نظریه کد بررسی می شود که چگونه مولکول خود را از خطا در برابر جهش[۱] محافظت می‌کند. در فصل اول این پایان‌نامه، بعضی مفاهیم و قضایا از نظریه کد، جبر و زیست‌شناسی ، که در فصل‌های بعدی مورد استفاده قرار می‌گیرند، را بیان می‌کنیم. مطالب این فصل برگرفته از مراجع [۵۳] و [۳۸] و [۴۶] می‌باشند.

۲.۱ تعاریف و مفاهیم لازم از جبر

در این بخش به بیان مفاهیم جبری لازم شامل گروه[۲]، حلقه[۳]، حلقه چندجمله‌ای‌های اریب[۴]، میدان[۵]، ... می‌پردازیم.

تعریف ۱.۲.۱. فرض کنیم $*$ یک عمل دوتایی روی مجموعه‌ی G باشد. در این صورت ساختمان جبری $(G,*)$ یک گروه است اگر شرایط زیر برقرار باشد:

الف) G نسبت به عمل $*$ بسته باشد.

ب) $*$ شرکت‌پذیر باشد.

ج) عضوی مانند $e \in G$ وجود داشته باشد به‌طوری که برای هر $x \in G$ داشته باشیم:

$$x * e = e * x = x.$$

د) برای هر $x \in G$، عنصر $x' \in G$ وجود داشته باشد به‌طوری که:

$$x * x' = x' * x = e.$$

اگر مجموعه‌ی G نسبت به عمل $*$ دارای خاصیت جابه‌جایی باشد؛ یعنی داشته باشیم:

$$\forall x, y \in G; \quad x * y = y * x;$$

در این‌صورت گروه $(G,*)$ گروه آبلی[۶] نامیده می‌شود.

[1] Mutation
[2] Group
[3] Ring
[4] Skew polynomial ring
[5] Field
[6] Abelian

عنصر e را عنصر همانی (خنثی) گروه و x' را وارون x می‌نامیم.

تعریف ۲.۲.۱. فرض کنید G یک گروه باشد. در این صورت زیرمجموعه‌ی H از G یک نیم‌گروه[1] G است اگر همراه با عمل دوتایی تعریف شده در G خود یک گروه باشد.

تعریف ۳.۲.۱. فرض کنیم R یک مجموعه باشد و دو عمل دوتایی با نام‌های جمع و ضرب روی R داشته باشیم. در این صورت R همراه با این دو عمل یک حلقه است اگر:

الف) R همراه با عمل جمع یک گروه آبلی باشد.

ب) عمل جمع شرکت‌پذیر باشد.

ج) عمل ضرب نسبت به جمع توزیع‌پذیر باشد، به‌عبارت دیگر برای هر $a,b,c \in R$ داریم:

$$a(b+c) = ab + ac \quad \text{و} \quad (a+b)c = ac + bc.$$

بنا به ۳.۲.۱ تعریف حلقه، ساختمان $(R,+)$ یک گروه آبلی است و در نتیجه یک عنصر خنثی دارد. در اینجا عنصر خنثی را با نماد $\circ$ نمایش خواهیم داد. بنابراین برای هر $a \in R$ داریم $a + \circ = \circ + a = a$.

د) عمل ضرب جابه‌جایی باشد. به‌عبارت دیگر برای هر $a,b \in R$ داشته باشیم $ab = ba$.

همچنین اگرعلاوه بر ۳.۲.۱ شرط زیر نیز برقرار باشد؛ آنگاه گوییم که R یک حلقه‌ی با عنصر همانی (حلقه‌ی یک‌دار) است.

ه) R نسبت به عمل ضرب عنصر همانی داشته باشد. در اینجا عنصر همانی R نسبت به عمل ضرب را با نماد ۱ نمایش می دهیم و آن را عنصر یکانی R می‌نامیم. بنابراین برای هر $a \in R$ داریم $a1 = 1a = a$.

تعریف ۴.۲.۱. فرض کنید R یک حلقه باشد. در این‌صورت زیرمجموعه S از R یک زیرحلقه[2] R است اگر:

الف) S تحت عمل جمع و ضرب R بسته باشد.

[1] Subgroup [2] Subring

ب) S با عمل جمع و ضرب القا شده از R خود یک حلقه باشد.

تعریف ۵.۲.۱. فرض کنیم R یک حلقه باشد و $a \in R$. در این صورت:

الف) a را یک مقسوم علیه صفر چپ می‌نامیم اگر عنصر $b \neq 0$ در R وجود داشته باشد به‌طوری که داشته باشیم $ab = 0$.

ب) a را یک مقسوم‌علیه صفر راست می‌نامیم اگر عنصر $c \neq 0$ در R وجود داشته باشد به‌طوری که داشته باشیم $ca = 0$.

پ) a را یک مقسوم‌علیه صفر[1] می‌نامیم اگر a یک مقسوم‌علیه صفر چپ یا مقسوم‌علیه صفر راست باشد.

تعریف ۶.۲.۱. یک دامنه صحیح یک حلقه‌ی جابجایی و یکدار است که فاقد مقسوم‌علیه صفر باشد.

تعریف ۷.۲.۱. زیرمجموعه‌ی I ازحلقه‌ی R یک ایده‌آل[2] است اگر:

۱. I یک زیرگروه جمعی R باشد.

۲. برای هر $a \in R$ و $i \in I$ داشته باشیم $ai, ia \in I$.

- I را یک ایده‌آل چپ[3] R گویند اگر به‌ازای هر $a \in R$ و هر $i \in I$ داشته باشیم $ai \in I$.
- I را یک ایده‌آل راست[4] R گویند اگر به‌ازای هر $a \in R$ و هر $i \in I$ داشته باشیم $ia \in I$ باشد.
- I را یک ایده‌آل دوطرفه‌ی R می‌گویند، اگر هم یک ایده‌آل چپ و هم یک ایده‌آل راست از R باشد.

تعریف ۸.۲.۱. فرض کنید X زیرمجموعه‌ای از حلقه R باشد. در این صورت اشتراک تمام ایده‌آل‌های R که شامل X باشند را ایده‌آل تولید شده توسط X نامند و آن را با نماد $\langle X \rangle$ نمایش می‌دهند.

[1] Zero divisor
[2] Ideal
[3] Left ideal
[4] Right ideal

تعریف ۹.۲.۱. عنصر x را مولد ایده‌آل X می‌نامند هرگاه $X = \langle x \rangle$ و ایده‌آل $\langle x \rangle$ را ایده‌آل اصلی تولید شده توسط x می‌نامند.

تعریف ۱۰.۲.۱. یک حلقه‌ی ایده‌آل اصلی حلقه‌ای است که هر ایده‌آل آن اصلی باشد. یک حلقه‌ی ایده‌آل اصلی که یک دامنه صحیح باشد، یک دامنه‌ی ایده‌آل اصلی نام دارد.

اگر I یک ایده‌آل از حلقه‌ی R باشد، در این صورت مجموعه‌ی $\frac{R}{I}$ را به صورت زیر تعریف می‌کنیم:

$$\frac{R}{I} = \{x + I | x \in R\}.$$

روی مجموعه $\frac{R}{I}$ می‌توان اعمال جمع و ضرب را به صورت زیر تعریف کرد:

$$(x + I) + (\acute{x} + I) = (x + \acute{x}) + I$$

$$(x + I).(\acute{x} + I) = (x\acute{x}) + I$$

می‌توان دید که مجموعه $\frac{R}{I}$ به همراه اعمال جمع وضرب فوق یک حلقه می‌باشد که آن را حلقه‌ی خارج قسمتی R بر I می‌نامند.

تعریف ۱۱.۲.۱. یک ایده‌آل M در یک حلقه R را ماکزیمال گویند اگر $M \neq R$ و به ازای هر ایده‌آل N در R اگر $M \subseteq N \subseteq R$ آن‌گاه داشته باشیم $N = R$ یا $N = M$.

تعریف ۱۲.۲.۱. فرض کنید $(R, +, .)$ و $(R', \oplus, \circ)$ دو حلقه باشند و $f : R \to R'$ یک تابع باشد. در این صورت f را یک همریختی[1] از R به R' گویند اگر به‌ازای هر $a, b \in R$ داشته باشیم:

الف) $f(a + b) = f(a) \oplus f(b)$.

ب) $f(a.b) = f(a)f(b)$.

تعریف ۱۳.۲.۱. فرض کنیم $f : R \to R'$ یک همریختی از حلقه R به حلقه‌ی R' باشد. در این صورت:

الف) $f : R \to R'$ را یک تکریختی (حلقه‌ای) نامند اگر f یک‌به‌یک باشد.

[1] Homomorphism

ب) f را یک برووریختی (حلقه‌ای) می‌باشد اگر f پوشا باشد.

ج) $f : R \to R'$ یک یکریختی (حلقه‌ای) می‌باشد اگر f هم یک‌به‌یک و هم پوشا باشد. اگر $f : R \to R'$ یک یکریختی (حلقه‌ای) باشد، گوییم که حلقه های R' و R یکریخت می‌باشند و می‌نویسیم $R \cong R'$.

تعریف ۱.۲.۱۴. فرض کنید R یک حلقه باشد.

الف) هر همریختی $f : R \to R$ را یک درونریختی از R می‌نامند.

ب) هر یکریختی $f : R \to R$ را یک خودریختی[1] از R می‌نامند.

تعریف ۱.۲.۱۵. فرض کنید $(R, +, .)$ یک حلقه باشد. کوچکترین عدد صحیح مثبت m به‌طوری که به‌ازای هر $a \in R$ داشته باشیم $ma = 0$ مشخصه‌ی حلقه‌ی R گویند و اگر چنین عدد صحیح مثبتی وجود نداشته باشد، گوییم R دارای مشخصه صفر است.

تعریف ۱.۲.۱۶. فرض کنید R یک حلقه باشد. یک دنباله‌ی نامتناهی $\alpha = (a_0, a_1, \dots, a_n, \dots)$ از عناصر R که در آن فقط تعداد متناهی از جملات غیرصفرند، را یک چندجمله‌ای[2] بر روی R می‌نامند. مجموعه‌ی تمام چندجمله‌ای‌های روی R را با $R[x]$ نشان می‌دهیم.

تعریف ۱.۲.۱۷. فرض کنید R یک حلقه باشد. عمل جمع و ضرب را روی $R[x]$ چنین تعریف می‌کنیم: به‌ازای هر

$$\alpha = (a_0, a_1, \dots, a_n, \dots)$$

و

$$\beta = (b_0, b_1, \dots, b_n, \dots)$$

داریم:

$$\alpha + \beta = (a_0 + b_0, a_1 + b_1, \dots, a_n + b_n, \dots)$$

و

$$\alpha.\beta = (c_0, c_1, \dots, c_n, \dots)$$

که در آن به ازای هر $k \in \mathbb{N} \cup \{0\}$ داریم $c_k = \sum_{i+j=k} a_i b_j$ می‌شود.

[1] Automorphism [2] Polynomial

تعریف ۱.۲.۱۸. اگر $\alpha \in R[x]$، در این صورت درجه α را با نماد $\deg(\alpha)$ نمایش می‌دهیم.

تعریف ۱.۲.۱۹. فرض کنید R یک حلقه‌ی جابجایی باشد و $a, b \in R$ باشند. گوییم که b بر a بخش‌پذیر است و یا a عاد می‌کند b را، و می‌نویسیم $a|b$، اگر عنصری مانند $x \in R$ وجود داشته باشد که به‌قسمی که $ax = b$. عناصر a و b از R را وابسته گویند، اگر $a|b$ و $b|a$.

تعریف ۱.۲.۲۰. فرض کنید R یک حلقه‌ی جابجایی و یکدار باشد.

الف) یک عنصر $c \in R$ را تحویل‌ناپذیر گویند اگر دو شرط زیر برقرار باشد:

۱) c یک عنصر غیرصفر غیریکه باشد.

۲) اگر $c = ab$ جایی‌که $ab \in R$ آن‌گاه، a یا b یک عنصر یکه است.

ب) یک عنصر $p \in R$ را اول[1] گویند اگر دو شرط زیر برقرار باشد:

۱) p یک عنصر غیرصفر یکه باشد.

۲) $p|ab$ و $a, b \in R$ نتیجه دهد $p|a$ یا $p|b$.

تعریف ۱.۲.۲۱. فرض کنید $\mathbb{N}_0$ مجموعه‌ی اعداد صحیح نامنفی باشد و R یک حلقه‌ی جابجایی باشد.

الف) R را یک حلقه‌ی اقلیدسی[2] می‌نامند، اگر یک تابع $\phi : R \to \mathbb{N}_0$ وجود داشته باشد به‌قسمی که شرایط زیر برقرار باشد:

۱) $\phi(n) = 0$ اگر و فقط اگر $a = 0$.

۲) به‌ازای هر $a, b \in R$ داریم $\phi(a, b) = \phi(a)\phi(b)$.

۳) به‌ازای هر $a, b \in R$، به‌قسمی که $b \neq 0$، عناصر $q, r \in R$ وجود دارند به‌طوری که:

$$a = bq + r \quad \phi(r) < \phi(b).$$

ب) یک حلقه‌ی اقلیدسی که یک دامنه‌ی صحیح باشد را دامنه‌ی اقلیدسی می‌نامند.

[1] Prime [2] Euclidean ring

تعریف ۲۲.۲.۱. مجموعه‌ی ناتهی F را همراه با دو عمل دوتایی جمع $(+)$ و ضرب $(.)$ میدان می‌نامیم هرگاه در شرایط زیر صدق کند:

۱. F نسبت به عمل جمع، بسته و یک گروه آبلی باشد. عنصر همانی نسبت به عمل جمع را عنصر صفر F می‌نامیم و با ۰ نمایش می‌دهیم.

۲. مجموعه عناصر ناصفر F تحت عمل ضرب بسته و یک گروه آبلی باشد. عنصر همانی نسبت به عمل ضرب را با ۱ نمایش می‌دهیم.

۳. عمل ضرب نسبت به عمل جمع توزیع‌پذیر باشد، یعنی:

$$a.(b+c) = a.b + a.c.$$

تعریف ۲۳.۲.۱. تعداد عناصر میدان را مرتبه[1] میدان گوییم. یک میدان با تعداد عناصر متناهی را یک میدان متناهی می‌نامیم.

تعریف ۲۴.۲.۱. فرض کنیم F یک میدان باشد. زیرمجموعه‌ی K از F که تحت اعمال F خود یک میدان باشد را زیرمیدان F می‌نامند. همچنین F را توسیع میدان K گویند.

تعریف ۲۵.۲.۱. فرض کنید F یک میدان باشد و $f(x) \in F[x]$. یک عنصر $s \in F$ را یک ریشه از چندجمله‌ای $f(x)$ یا از f گویند، اگر $f(s) = ۰$. در این‌صورت می‌گویند s در معادله‌ی چندجمله‌ای $f(x)$ صدق می‌کند.

تعریف ۲۶.۲.۱. دو عمل روی F_q^n تعریف می‌کنیم:

الف) جمع دو بردار: برای هر $V = (v_۰, v_۱, \ldots, v_{n-۱})$ و $U = (u_۰, u_۱, \ldots, u_{n-۱})$ در F_q^n بردار $U+V$ به فرم زیر می‌باشد:

$$U + V = (u_۰ + v_۰, u_۱ + v_۱, \ldots, u_{n-۱} + v_{n-۱}) \in F_q^n.$$

ب) حاصل‌ضرب یک بردار در یک اسکالر: اگر $aV = (v_۰, v_۱, \ldots, v_{n-۱}) \in F_q^n$ و $a \in F_q$ باشد، آن‌گاه داریم $aV = a(v_۰, v_۱, \ldots, v_{n-۱}) = (av_۰, av_۱, \ldots, av_{n-۱}) \in F_q^n$.

[1] Order

مجموعه‌ی $\mathbb{F}_q^n$ به همراه عمل جمع و ضرب اسکالر فوق یک فضای برداری روی میدان $\mathbb{F}_q$ می‌باشد.

تعریف ۲۷.۲.۱. یک زیرمجموعه از F_q^n یک زیرفضا[1] از F_q^n است هرگاه تحت عمل جمع و ضرب تعریف شده روی F_q^n یک فضای برداری باشد.

تعریف ۲۸.۲.۱. فرض کنید V یک فضای برداری روی F_q باشد و $S = \{v_1, v_2, \ldots, v_k\}$ زیرمجموعه‌ای ناتهی از V باشد. در این‌صورت زیرفضای تولیدشده توسط S تعریف می‌شود:

$$\langle S\rangle = \{\lambda_1 v_1 + \cdots + \lambda_k v_k | \forall_i; \lambda_i \in F_q\}$$

اگر $S = \circ$ باشد، تعریف می‌کنیم $\langle S\rangle = \{o\}$.

تعریف ۲۹.۲.۱. ماشین تورینگ قطعی[2] $\delta : \Gamma \times \Sigma \to \Gamma \times \Sigma \times \{\leftarrow, \rightarrow\}$

الف) مجموعه متناهی Σ که الفبا نامیده می‌شود و شامل کاراکتر $\star$ است که به‌معنی جای خالی می‌باشد.

ب) یک نوار دو طرفه نامتناهی که به مربع‌هایی تقسیم شده است. در همه مربع‌های این نوار $\star$ قرار گرفته است به جز تعداد متناهی از آن‌ها که با اعضای دیگری از Σ پر شده‌اند. هر مربع شامل $\star$ به معنی یک مربع خالی است.

ج) سوزن خواندن و نوشتن[3] این سوزن در هر لحظه می‌تواند مقدار مربعی که روی آن قرار گرفته است را بخواند و سپس مقدار آن را در صورت لزوم تغییر دهد و بعد از آن به سمت راست یا چپ نوار حرکت کند.

د) واحد کنترل کننده[4] با یک مجموعه متناهی از حالت‌های ممکن Γ که $\Gamma : \{\gamma_\circ, \gamma_1, \ldots, \gamma_t\}$ که ابتدا در حالت $\gamma_\circ$است؛ هم‌چنین Γ شامل زیرمجموعه از حالت‌هایی است که ماشین در آن حالت‌ها متوقف می‌شود. نحوه کار ماشین تورینگ به کمک تابع انتقال مشخص می‌شود.

$$\delta : \Gamma \times \Sigma \to \Gamma \times \Sigma \{\leftarrow, \rightarrow\}$$

[1] Subspace
[2] Deterministic Turing Machines
[3] Read-write head
[4] Control unit

تعریف ۳۰.۲.۱. زبان $L \subseteq \sum_{\circ}$ را متعلق به NP گوئیم، هرگاه یک ماشین تورینگ قطعی M و یک چندجمله‌ای $P(n)$ چنان وجود داشته باشد که اولاً $T_M(n) \leq p(n)$ و همچنین برای هر $x \in \sum_{\circ}$ داشته باشیم:

- اگر $x \in L$، آن‌گاه یک تصدیق $y \in \sum_{\circ}$ چنان وجود داشته باشد که $|y| \leq p(n)$ و ماشین M ورودی $x - y$ را تصدیق کند.
- اگر $x \notin L$ ، آن‌گاه برای هر رشته $y \in \sum_{\circ}$ و خروجی M روی ورودی $x - y$ ، NO باشد.

L را NP کامل گوییم هرگاه:

- $L \in NP$
- $\grave{L} \in NP; \cdot \grave{L} \leq L$

فرض کنید E یک زیرفضای خطی K بعدی $\mathbb{R}^n$ به صورت $E = \langle A^1, \ldots, A^l \rangle$ که $\left\{A^1, \ldots, A^l\right\}$ مستقل خطی می‌باشند همچنین هر دو بردار که مستقل‌خطی باشند پایه تشکیل می‌دهند؛ حال می‌خواهیم با استفاده از **روش گرام‌ـ اشمیت**[1] پایه‌های یکا متعامد را بدست آوریم که بردارها دو به دو بر هم عمود و طول آن‌ها یک باشند.

A^1 را تثبیت می‌کنیم داریم $A^1 = B^1$ تصویر قائم را کم می‌کنیم و طبق ضرب داخلی داریم $B^2 = A^2 - \frac{A^2.B^1}{|B^1||B^1|}.B^1$

ادعا می‌کنیم $B^2 \perp B^1$ و ترکیبات خطی $\langle B^1, B^2 \rangle = \langle A^1, A^2 \rangle$ بررسی می‌کنیم ضرب داخلی آن‌ها صفر شود لذا داریم

$$B^2.B^1 = A^2.B^1 - \frac{A^2.B^1}{|B^1||B^1|}B^1.B^2 = \circ$$

پس $B^1, B^2 \in \langle A^1, A^2 \rangle$ لذا هرگاه دو بردار داشته باشیم که مستقل خطی باشند پایه می‌شود و فضا را ایجاد می‌کنند و چون B_1, B_2 برهم عمود می‌باشند B^1 همان A^1 و B^2 ناصفر می‌باشد. بردارهایی که دو به دو بر هم عمود باشند را بدست می آوریم؛ سپس هر عضو پایه را در معکوس طول آن ضرب می‌کنیم و پایه های متعامد یکا بدست می‌آیند.

[1] Gram-Schmidt

۳.۱ تعاریف و مفاهیم لازم از نظریه کدگذاری

تعریف ۱.۳.۱. فرض کنید A یک مجموعهٔ qـتایی به‌صورت $A = \{a_1, a_2, \ldots, a_q\}$ باشد. مجموعهٔ A الفبای کد نامیده می‌شود و هر یک از عناصر مجموعهٔ A را *سمبل (نماد)* کد[1] گویند. همچنین فرض کنید A^n تمامی رشته‌های به‌طول n روی مجموعهٔ A باشد؛ یعنی:

$$A^n = \{x_1 x_2 \ldots x_n \mid x_i \in A\}.$$

در این صورت:

الف) هر زیرمجموعهٔ ناتهی مانند C از A^n یک کد بلوکی qـتایی نامیده می‌شود.

ب) هر عضو C یک کدواژه[2] نامیده می‌شود.

ج) تعداد کدواژه‌های کد C را *اندازهٔ* کد نامیده و با نماد $|C|$ نمایش داده می‌شود.

د) یک کد از طول n و اندازهٔ M را یک (n, M)ـکد گویند.

گاهی اوقات، کد بلوکی qـتایی C را کد qـتایی یا به‌طور خلاصه‌تر کد می‌گویند.

ملاحظه ۱.۳.۱. اکثر مواقع الفبای کد یک میدان متناهی در نظر گرفته می‌شود؛ در این صورت $A = F_q$ می‌باشد.

- اگر $A = F_2$ باشد؛ در این صورت کد را یک کد دودویی[3] می‌نامند.
- اگر $A = F_3$ باشد؛ در این صورت کد را یک کد سه‌تایی[4] می‌نامند.
- اگر $A = F_4$ باشد؛ در این صورت کد را یک کد چهارتایی[5] می‌نامند.

تعریف ۲.۳.۱. فرض کنید a و b دو کلمه به‌طول n روی الفبای A باشند. فاصله همینگ بین دو کدواژهٔ $a = (a_1, \ldots, a_n)$ و $b = (b_1, \ldots, b_n)$ که با $d_H(a, b)$ نمایش داده می‌شود، به صورت تعداد جایگاه‌های متفاوت a و b تعریف می‌شود؛ یعنی تعداد بیت‌هایی که a و b با هم تفاوت دارند. لذا داریم:

$$d_H(a, b) = \Big|\{i \mid a_i \neq b_i \ ; \ i = 1, \ldots, n\}\Big|.$$

[1] Code symbol
[2] Codeword
[3] Binary code
[4] Ternary code
[5] Quaternary code

به‌بیان معادل $d_H(a,b) = \sum_{i=1}^{n} d_H(a_i,b_i)$ می‌باشد؛ جایی که،

$$d_H(a_i,b_i) = \begin{cases} 1 & ; a_i \neq b_i \\ 0 & ; a_i = b_i. \end{cases}$$

تعریف ۳.۳.۱. فرض کنید C یک کد بلوکی دارای حداقل دو کدواژه باشد. در این صورت فاصلهٔ کد[1] C که با نماد $d(C)$ نمایش داده می‌شود، به صورت مینیمم فاصله همینگ بین تمام کدواژه‌های کد C تعریف می‌شود، یعنی داریم:

$$d(C) = min\{d_H(a,b) \mid a,b \in C \ ; \ a \neq b\}.$$

ملاحظه ۲.۳.۱. هر کد بلوکی که طول آن n (طول کدواژه‌های کد بلوکی)، اندازهٔ آن M (تعداد کدواژه‌های کد C) و فاصله همینگ آن d باشد را یک (n,M,d)ـ کد می‌نامند. در این صورت اعداد n، M و d را پارامترهای کد گویند.

یک کد خطی[2] از طول n روی میدان متناهی F_q، یک زیرفضایی از فضای برداری F_q^n می‌باشد. در این حالت اگر C دارای بُعد k باشد، آن را به‌صورت $[n,k]$ـ کد نمایش می‌دهند و اگر $d(C) = d$ فاصلهٔ کد C باشد؛ آن‌گاه C را یک $[n,k,d]$ـ کد خطی گویند.
لازم به ذکر است که اگر کد C خطی نباشد، آن را غیرخطی می‌نامند.

تعریف ۴.۳.۱. فرض کنید $x = x_1,\ldots,x_n$ یک کلمه در F_q^n باشد. در این‌صورت وزن همینگ[3] کلمهٔ x که با نماد $wt(x)$ نمایش داده می‌شود، برابر با تعداد مؤلفه‌های ناصفر کلمهٔ x تعریف می‌شود؛ لذا داریم: $wt(x) = d(x,0)$، جایی که $0 = (0,0,\ldots,0) \in F_q^n$. به‌عبارت دیگر، $wt(x) = \sum_{i=1}^{n} wt(x_i)$ جایی که،

$$wt(x_i) = d(x_i,0) = \begin{cases} 1 & ; x_i \neq 0 \\ 0 & ; x_i = 0. \end{cases}$$

لم ۱.۳.۱. اگر x و y دو عنصر دلخواه در $\mathbb{F}_q^n$ باشند، در این‌صورت همواره داریم:

$$d(x,y) = wt(x-y).$$

[1] Distance of a code
[2] Linear code
[3] Hamming weight

برهان. ابتدا توجه داریم که برای هر $x = (x_۱, \ldots, x_n) \in \mathbb{F}_q^n$ و $y = (y_۱, \ldots, y_n) \in \mathbb{F}_q^n$ همواره داریم:

$$wt(x) = \sum_{i=۱}^{n} wt(x_i) = wt(x_۱) + wt(x_۲) + \cdots + wt(x_n)$$
$$d(x, y) = \sum_{i=۱}^{n} d(x_i, y_i) = d(x_۱, y_۱) + d(x_۲, y_۲) + \cdots + d(x_n, y_n)$$

حال کافی است نشان دهیم برای $x_i \in \mathbb{F}_q$ و $y_i \in \mathbb{F}_q$، همواره داریم: $d(x_i, y_i) = wt(x_i - y_i)$ که این مطلب واضح است زیرا:

$$d(x_i, y_i) = ۰ \iff x_i = y_i \iff x_i - y_i = ۰ \iff wt(x_i - y_i) = ۰$$

□

۱.۳.۱ قواعد کدگشایی

در این قسمت دو قاعده‌ی کدگشایی که برای کدها مورد استفاده قرار می‌گیرند را بیان می‌کنیم.

۲.۳.۱ قاعده کدگشایی ماکزیمم احتمال (MLD)

فرض کنید کدواژه‌های کد C از طریق یک کانال پارازیت‌دار ارسال شده باشند و کلمه x را دریافت کرده باشیم. در این صورت قاعده کدگشایی ماکزیمم احتمال[۱] ، کلمه x را به کدواژه c_x کدگشایی می‌کند هرگاه احتمال ارسال برای c_x به شرط دریافت x ماکزیمم باشد. دو نوع قاعده کدگشایی ماکزیمم احتمال داریم:

۱. قاعده کدگشایی ماکزیمم احتمال کامل (CMLD)

دراین روش اگر کلمه x دریافت شود و تعداد کدواژه‌هایی که به شرط دریافت کلمه x دارای ماکزیمم احتمال ارسال می‌باشند، بیش از یک مورد باشد، آن‌گاه یکی از کدواژه‌ها را به دلخواه انتخاب کرده و کلمه x را به آن کدواژه کدگشایی می‌کند.

۲. قاعده کدگشایی ماکزیمم احتمال غیرکامل (IMLD)

دراین روش اگر کلمه x دریافت شود و تعداد کدواژه‌هایی که به شرط دریافت کلمه x

[1] Maximum likelihood decoding

دارای ماکزیمم احتمال ارسال است، بیش از یک مورد باشد، آن‌گاه بدون هیچ‌گونه تصمیم‌گیری صرفاً درخواست ارسال مجدد را انجام می‌دهد.

۳.۳.۱ قاعده کدگشایی مینیمم فاصله (MDD)

فرض کنید کدواژه‌های کد C از طریق یک کانال پارازیت‌دار ارسال شده باشند و کلمه x را دریافت کرده باشیم؛ دراین صورت قاعده کدگشایی مینیمم فاصله[1] ، کلمه x را به کدواژه c_x کدگشایی می‌کند هرگاه فاصله بین x و c_x در بین تمام کدواژه‌های C ، مینیمم مقدار ممکن باشد؛ یعنی داشته باشیم:

$$d_H(x, c_x) = \min \{d_H(x, c) | c \in C\}.$$

دو نوع قاعده کدگشایی مینیمم فاصله داریم:

۱. قاعده کدگشایی مینیمم فاصله کامل (CMDD)

در این روش اگر کلمه x دریافت شود و تعداد کدواژه‌هایی که با کلمه x دارای مینیمم فاصله می‌باشند، بیش از یک مورد باشد؛ آن‌گاه یکی از کدواژه‌ها را به‌دلخواه انتخاب کرده و کلمه x را به آن کدواژه کدگشایی می‌کند.

۲. قاعده کدگشایی مینیمم فاصله غیرکامل (IMDD)

در این روش اگر کلمه x دریافت شود و تعداد کدواژه‌هایی که با کلمه x دارای مینیمم فاصله است، بیش از یک مورد باشد؛ آن‌گاه بدون هیچ‌گونه تصمیم‌گیری صرفاً درخواست ارسال مجدد را انجام می‌دهد.

تعریف ۵.۳.۱. کد C، t ــ تشخیص‌دهنده خطا نامیده می‌شود، حداقل یک خطا و حداکثر t خطا روی هر کدواژه رخ دهد؛ آن‌گاه کلمه دریافتی، یک کدواژه از کد C نباشد. همچنین کد C یک کد دقیقاً تشخیص‌دهنده t خطا نامیده می‌شود، هرگاه تشخیص‌دهنده t خطا باشد؛ اما تشخیص‌دهنده $t+1$ خطا نباشد.

می‌توان دید که کد C ، یک کد t ــ تشخیص دهنده خطاست اگر و تنها اگر $d(C) \geq t+1$ باشد. و هر کد C با فاصله $d(C) = d$ ، یک کد دقیقاً تشخیص‌دهنده $d-1$ خطا باشد.

[1] Minimum distance decoding

تعریف ۱.۳.۶. کد C، t ـ تصحیح‌کننده خطا نامیده می‌شود، اگر قاعده (MDD) قادر به تصحیح حداکثر t خطا باشد. همچنین کد C دقیقاً تصحیح‌کننده t خطاست اگر تصحیح کننده t خطا باشد؛ اما تصحیح‌کننده $t+1$ خطا نباشد.

می توان دید که کد C یک کد تصحیح‌کننده t خطاست اگر و تنها اگر $d(C) \geq 2t+1$ باشد. فرض کنید کد C یک کد با فاصله $d(C) = d$ باشد؛ در این‌صورت کد C دقیقاً تصحیح‌کننده $\frac{d-1}{2}$ خطا می‌باشد.

۱.۳.۴ کران‌هایی در نظریه کدگذاری

فرض کنید C یک (n, M, d) کد qـتایی باشد. در این‌صورت کارایی کد با مقادیر M و d در ارتباط است. در واقع ، یک (n, M, d) ـ کد خوب باید دارای دو شرط زیر باشد:

الف) دارای اندازه M بزرگ باشد تا بتواند تعداد بیشتری از پیام‌های منبع را کد نماید.

ب) دارای بیشترین فاصله d ممکن باشد تا بتواند خطاهای بیشتری را تصحیح کند.

تعریف ۱.۳.۷. فرض کنید مجموعه الفبای A مفروض باشد به‌طوری که $|A| = q \geq 1$ و مقادیر d و n مشخص باشند. در این‌صورت $A_q(n, d)$ را ماکزیمم اندازه ممکن برای M ؛ به‌گونه‌ای که یک (n, M, d) کد موجود باشد، تعریف می‌کنند.

ملاحظه ۱.۳.۳. محاسبه مقدار $A_q(n, d)$ مساله‌ای جالب و تاثیرگذار در نظریه کدگذاری می‌باشد که از آن به‌عنوان مهم‌ترین مسئله نظریه کدگذاری یاد می‌شود. هر چند محاسبه مقدار دقیق $A_q(n, d)$ کار ساده‌ای نمی‌باشد؛ ولی می‌توان برای آن کران‌هایی بالا و پایین مشخص کرد که در عمل، کار را راحت می‌کند.

قضیه ۱.۳.۱. کران همینگ (کران گوی‌بندی):
برای عدد صحیح $q > 1$ و اعداد صحیح n و d به گونه‌ای که $1 \leq d \leq n$، همواره داریم

$$A_q(n, d) \leq \frac{q^n}{\sum_{i=0}^{\lfloor \frac{d-1}{2} \rfloor} \binom{n}{i}(g-1)^i}.$$

برهان. قضیه ۱.۳.۵ مرجع [۲۹] را ببینید. □

تعریف ۸.۳.۱. هر کد qـتایی که در شرایط مرزی کران همینگ صدق کند، یعنی دارای $\frac{q^n}{\sum_{i=0}^{\lfloor \frac{d-1}{2} \rfloor} \binom{n}{i}(g-1)^i}$ کدواژه باشد، یک کد کامل[1] نامیده می‌شود.
کدهای همینگ یک رده مهم از کدهای کامل می‌باشند که به‌صورت زیر تعریف می‌شوند.

تعریف ۹.۳.۱. فرض کنیم $r \geq 2$ باشد. یک کد خطی دودویی از طول $n = 2^r - 1$، که ستون‌های ماتریس کنترل توازن آن متشکل از تمام بردارهای ناصفر $\mathbb{F}_2^r$ می‌باشد،کد همینگ دودویی[2] از طول $2^r - 1$ نامیده می‌شود و با نماد $\text{Ham}\ (r,2)$ نمایش داده می‌شود.

مثال ۱.۳.۱. کد همینگ $\text{Ham}\ (3,2)$، از طول ۷ و با ماتریس کنترل توازن زیر می‌باشد.

$$H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

گزاره ۱.۳.۱. (برخی خواص کدهای همینگ دودویی)

الف) تمام کدهای همینگ دودویی از یک طول مشخص، هم ارز می‌باشند.

ب) بعد کد همینگ $\text{Ham}\ (r,2)$ برابر است با $K = 2^r - 1 - r$.

ج) فاصله کد $\text{Ham}\ (r,2)$ برابر ۳ می‌باشد، لذا $\text{Ham}\ (r,2)$ یک کد دقیقاً تصحیح‌گر ۱ـخطا می‌باشد.

د) کدهای همینگ دودویی، کدهایی کامل می‌باشند.

برهان. گزاره ۶.۳.۵ مرجع [۲۹] را ببینید. □

تعریف ۱۰.۳.۱. فرض کنیم $r \geq 2$ باشد. یک کدخطی q تایی، که ستون‌های ماتریس کنترل توازن آن متشکل از دقیقاً یک بردار ناصفر از هر زیر فضای برداری با بعد یک از $\mathbb{F}_q^r$ می‌باشد، یک کد همینگ qـ تایی نامیده می‌شود و با نماد $\text{Ham}\ (r,2)$ نمایش داده می‌شود.

[1] Perfect code
[2] Binary Hamming code

۴.۱ تعاریف و مفاهیم لازم از زیست شناسی سلولی و مولکولی

در این قسمت تعاریف و مفاهیم لازم از زیست‌شناسی سلولی و مولکولی که در ادامه از آن‌ها استفاده خواهیم کرد را بیان می کنیم.

تعریف ۱.۴.۱. کوچکترین واحد حیاتی پیکره هر موجود زنده را سلول[۱] می‌نامند. چنانچه بدن موجود زنده را به یک ساختمان تشبیه کنیم، سلول‌ها به تعبیری آجرهای تشکیل‌دهنده بنا هستند.

هر سلول از اجزای مختلفی تشکیل شده است. اما سه بخش اصلی تشکیل‌دهنده آن عبارتند از:

الف) هسته[۲]

ب) سیتوپلاسم[۳]

ج) غشای سلول[۴] .

تعریف ۲.۴.۱. سیتوپلاسم ماده‌ای است که فضای درون سلول را پر می‌کند. ترکیب اصلی سیتوپلاسم آب و پروتئین محلول در آب است.

تعریف ۳.۴.۱. هسته اندامکی است که در اکثر سلول‌های یوکاریوت[۵] (پرسلولی‌ها) وجود دارد و بیشتر حاوی اطلاعات ژنتیکی است. هسته‌ی سلول دو وظیفه دارد:

الف) کنترل واکنش‌های شیمیایی درون سیتوپلاسم

ب) نگه داشتن اطلاعات لازم برای تقسیم سلول .

تعریف ۴.۴.۱. DNA نوعی اسید نوکلئیک[۶] (اسید نوکلئیک‌ها پلیمرهای خطی هستند که از تکرار واحدهایی به‌نام نوکلئوتید تشکیل شده‌اند)می باشد. در واقع نوکلئوتیدها واحدهای

[1] Cell
[2] Nucleus
[3] Cytoplasm
[4] Cell membrance
[5] Eukaryote
[6] Nucleic acid

تک‌پار اسیدهای نوکلئیک به‌شمار می‌آیند، که دارای پنج قند کربنه پنتوز، باز هتروسیکلیک و گروه فسفات می‌باشند

DNA دستورالعمل ژنتیکی مورد استفاده در توسعه و عملکرد تمام موجودات زنده شناخته‌شده را کدگذاری می‌کند و دو ـ رشته‌ای می‌باشد.

تعریف ۵.۴.۱. RNA[1] گروه دیگری از نوکلئیک اسیدها است که تک‌رشته‌ای و کوتاه‌تر از DNA می‌باشند.

تعریف ۶.۴.۱. ژن[2] واحد فیزیکی و عملکرد وراثت است. ژن‌ها از DNA ساخته شده‌اند. برخی ژن‌ها به عنوان راهنما برای ساختن مولکول‌هایی به نام پروتئین عمل می‌کنند. پروژه ژنوم انسانی تخمین زده است که انسان‌ها بین ۲۰۰۰۰ تا۲۵۰۰۰ ژن دارند. هر فرد دارای دو نسخه از هر ژن می‌باشد، که هر کدام را یکی از والدین به ارث برده است.

تعریف ۷.۴.۱. مجموعه ژن‌های یک سلول را ژنوم می‌گویند. همچنین یک مجموعه کروموزمی که به عنوان واحد از والدین به فرزندان ارث می‌رسد، ژنوم گفته می‌شود. برخی بیولوژیست‌ها تمام ماده ژنتیک موجود زنده حتی یک سلول را ژنوم می‌نامند.

تعریف ۸.۴.۱. نوکلئوتیدها[3] واحدهای اسیدنوکلئیک به‌حساب می‌آیند. نوکلئوتیدها نقش‌هایی را در فرایندهای مربوط به سیگنالینگ سلولی، متابولیسم و واکنش‌های آنزیمی بر عهده دارند. ساختار کلی آن شامل بازها آلی (انتقال پیام) پروتون پذیرنده معمولاً حاوی اتم نیتروژن هستند که قند پنج‌کربنی و دست‌کم یک گروه فسفات تشکیل شده است.

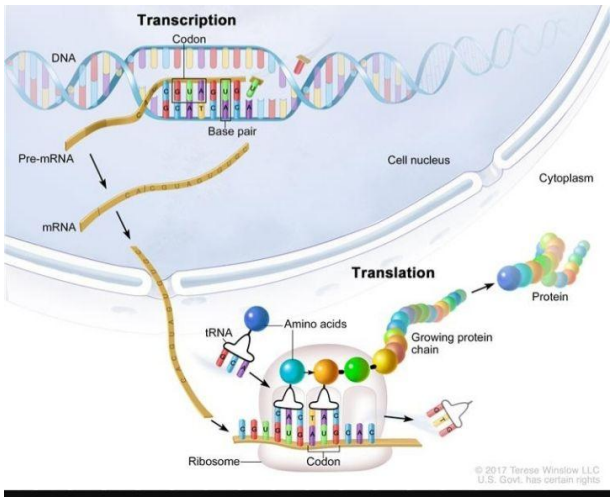
تعریف ۹.۴.۱. دو دسته از بازهای مهم و اصلی اسیدنوکلئیک، پورین‌ها[4] (A[5] G[6]) و پیریمیدین‌ها[7] (C[8] و T[9]، U[10]) می‌باشند.

DNA از چهار پایه یا نوکلئوتید تشکیل شده است. آدنین و گوانین و سیتوزین و تیمین. آدنین و گوانین تحت عنوان پورین در نظر گرفته می‌شوند و با R نمایش داده می‌شود (دوحلقه‌ای)، سیتوزین و تیمین به‌عنوان پیریمیدین (تک حلقه‌ای) Y هستند . پورین‌ها از پیریمیدین‌ها

[1] Ribonucleic acid
[2] Gene
[3] Nucleotid
[4] Purin
[5] A=Adenin
[6] G=Guvanin
[7] Prymidin
[8] C=Cytozin
[9] T=Tymin
[10] U=Uracyl

بزرگتر هستند و این اندازه و عدم تعادل بین نوکلئوتیدها میل ترکیبی بین پورین و پیریمیدین به‌وجود می‌آورد و ثبات آن‌ها تنها با جفت شدن ایجاد می‌شود. A به T وصل می‌شود (دو پیوند ضعیف هیدروژن) و C به G وصل می‌شود (سه پیوند ضعیف هیدروژن).

تعریف ۱.۴.۱۰. از روی الگوی DNA طی فرایندی به‌نام رونویسی[۱]، نسخه‌برداری می‌شود و رشته RNA به وجود می‌آید. هر DNA به‌عنوان الگوی سنتز در RNA به‌کار می‌رود.
ژن‌ها، در صورتی فعال می‌باشند که در حال نسخه‌برداری از DNA باشند و مولکول پروتئین طی فرایندی به‌نام ترجمه (از روی RNA پروتئین‌ها ساخته شود) سنتز می‌شود. پروتئین‌ها شامل آمینواسیدها می‌باشند. به‌طور معمول سه نوکلوتید مجاور هم یک آمینواسید را به‌وجود می‌آورند که کدون[۲] یا رمزینه نامیده می‌شود. فقط mRNA پیک حاوی رمز ژنتیک است. (شکل ۱.۱ رابببینید)



شکل ۱.۱: رونویسی و ترجمه در DNA

تعریف ۱.۴.۱۱. مدل واتسون و کریک[۳]: جیمز واتسون و کریک مدل مارپیچ دورشته‌ای را بیان کردند که هر مولکول DNA از دورشته پلی‌نوکلئوتید ساخته شده که به دور محوری فرضی می‌پیچد و ساختار مارپیچ دورشته‌ای را ایجاد می‌کند که پله‌های این مارپیچ بازهای آلی هستند که با پیوند هیدروژنی به هم متصل می‌باشند.

[1] Transcription
[2] Codon
[3] Watson-Crick

تعریف ۱.۴.۱۲. کروموزوم[1] یکی از واحدهای ژنوم است که چندین ژن را در بر می‌گیرد. هر کروموزوم از یک واحد مولکول DNA دو رشته‌ای طویل و تقریباً به همان مقدار پروتئین ساخته شده است. ساختمان نوکلئوپروتئینی در دو انتهای کروموزوم سلولی‌ها را اصلاحاً تلومر[2] می‌گویند. نقش تلومر به اختصار عبارت است از:

الف) پایداری کروموزوم

ب) استقرار کروموزوم

ج) سرکوب فرآیند رونویسی.

تعریف ۱.۴.۱۳. اگزونوکلئازها آنزیم‌هایی هستند، که از گروه نوکلئاز برای شکستن پیوند فسفو دی استر بین نوکلئوتیدها در دورشته DNA به کار می‌برند؛ همچنین آنزیم اگزونوکلئازها برای صاف کردن انتهای مولکول‌های DNA که دارای انتهای چسبنده می‌باشند، به کار گرفته می‌شوند.

تعریف ۱.۴.۱۴. تنها یک درصد از DNA از ژن‌های کدکننده پروتئین ساخته شده است؛ و ۹۹ درصد دیگر غیر کدکننده است. DNA غیر کدکننده دستورالعملی برای ساخت پروتئین ندارد. این DNA حاوی توالی‌هایی است که به عنوان عناصر تنظیم یعنی تعیین این که چه زمانی و کجا ژن‌ها خاموش و روشن شوند عمل می‌کند.

تعریف ۱.۴.۱۵. بخش عمده ناحیه دو رشته‌ایی تلومر DNA به‌وسیله روش نیمه‌حفاظتی و توسط DNA پلیمراز ساخته می‌شود. باقی‌مانده ناحیه دو رشته‌ای و همین‌طور بخش تک‌رشته تلومر توسط آنزیم تلومراز ساخته می‌شود.

تعریف ۱.۴.۱۶. شبکه تنظیم کننده ژن، مجموعه‌ای از قطعات DNA می‌باشند که با یکدیگر به‌صورت غیرمستقیم در ارتباط هستند. برخی از پروتئین‌ها تنها وظیفه فعال کردن ژن‌های دیگر را برعهده دارند و این عوامل رونویسی هستند که نقش اصلی در شبکه نظارتی را برعهده دارند.

[1] Chromosome [2] Telumer

تعریف ۱.۴.۱۷. جهش[1] در سلول به ۲ صورت اتفاق می‌افتد:

۱. ژنی (نقطه‌ای، حذفی، افزایشی)

۲. کشنده

جهش ژنی شامل انواع نقطه‌ای (Point Mutation)، افزایشی (Addition) و حذفی (Delation) است. در جهش نقطه‌ای یک جفت باز تغییر می‌کند. در جهش‌های حذفی و افزایشی یک یا بیش از یک جفت نوکلئوتید حذف یا اضافه می‌شود.

در جهش نقطه‌ای موسوم به Missense ویژگی کدون عوض می‌شود به طوری که آن کدون آمینواسید دیگری را کد می‌کند. این نوع جهش به دو گروه Transition و Transversion تقسیم می‌شوند، که در نوع اول نوکلئوتیدهای پورینی و پیریمیدنی با انواعی از همنوع خود عوض می‌شوند و در نوع دوم نوکلئوتیدهای پورین و پیریمیدین با همدیگر جایگزین می‌شوند. می‌توان به شکل مستطیلی مدل‌سازی کرد که اضلاع مستطیل معرف جهش Transition و دو قطر آن جهش Transversion را نشان می‌دهد. در جهش حذفی یا افزایشی اگر یک یا دو نوکلئوتید تغییر کند، ویژگی کدون عوض می‌شود که اصطلاحاً گفته می‌شود قاب خواندن تغییر کرده است و جهش تغییر قاب[2] اتفاق افتاده است. در صورتی که یک کدون سه حرفی حذف و یا یک توالی سه حرفی فاصله بین کدون‌ها اضافه شود، قالب خواندن تغییر نخواهد کرد.

[1] Mutation

[2] Frame Shift Mutation

فصل ۲

کدگذاری توالی DNA

۱.۲ مقدمه

امروزه علوم ریاضی و محاسبات از جایگاه خاصی در زیست‌شناسی برخوردار بوده و کاربردهای فراوانی در این زمینه دارد. تبادل پردازش اطلاعات قابل اطمینان، هم نیازمند سیستم مصنوعی و هم نیازمند سیستم زیستی می‌باشد. در پردازش اطلاعات مصنوعی، روش کدگذاری نقش مهمی برای تضمین اطمینان مورد نیاز بازی می‌کند. با پیشرفت فناوری، پژوهشگران توانستند داده‌های متنوع‌تری از موجودات زنده را به‌دست آورند. یکی از فعالیت‌های اخیر پژوهشگران، شناخت بیماری‌ها و مدل‌سازی آن‌ها به کمک ریاضیات می‌باشد که منجر به شناخت بهتر بیماری‌های دشوار درمان می‌شود. به‌طور مثال پیش‌بینی این‌که چند درصد انسان‌ها دچار سرطان می‌شوند، می‌تواند حائز اهمیت باشد. کارهای مختلفی برای توصیف و در نتیجه شناخت بهتر سرطان، انجام شد که به‌عنوان نمونه، مدل‌سازی ریاضی، به همراه معادله‌های ریاضی، برای پدیده رشد سلول‌های سرطانی می‌باشد. سال‌های گذشته میان دو گروه مخالف و موافق بحثی درباره تصادفی بودن پدیده سرطان در گرفت. افرادی معتقد بودند سرطان پدیده تصادفی است و کوشش نکنید زندگی سالمی داشته باشید و سرطان به سبک زندگی ربط چندانی ندارد؛ اما گروه مقابل معتقد بودند که سبک زندگی تاثیر مهمی در ابتلا به سرطان دارد. هر دو گروه هم مدل‌سازی‌های ریاضی انجام داده بودند که بتوانند با داده‌ها نتایج را توجیه نمایند. این نمونه از پژوهش‌ها که با پیشنهاد درمان فاصله زیادی دارند، ولی همین مدل‌سازی‌های ریاضی، تاثیر زیادی در شناخت درست بیماری‌ها دارند. جین مایرز[۲]

[2] Jane Myers

ریاضیدان دانشگاه آلیزونا با استفاده از الگوریتمی کامپیوتری توانست توالی قطعات DNA در ژنوم مگس سرکه را تعیین کند. قدم بعدی تعیین نقشه ژنوم انسان است. موارد ذکر شده نمونه‌هایی از پژوهش‌های اخیر در ریاضیات زیستی می‌باشد.

این فصل در مورد کاربردهای نظریه کدگذاری در زیست‌شناسی سلولی و مولکولی می‌باشد. با مدل‌سازی و الگوریتم‌های پیشنهادی، توالی تکراری DNA را بررسی می‌کنیم. مطالب این فصل برگرفته از مراجع [۵۳] و [۴۲] است.

۲.۲ نمایش نوکلئوتیدها

کد ژنتیکی[1] پروتئین‌هایی را از انتقال نوکلئوتیدها به آمینواسیدها می‌سازد. اما این نمونه یکی از هزاران پیامی است که در DNA کدگذاری می‌شود. کدگذاری پروتئین‌ها، کمترین سرعت میزان جهش و دگرگونی را در رشته DNA دارند. از زمان معرفی مدل واتسون ـ کریک، محققان در تلاشند تا حداکثر توالی میلیونی پایه را کدگشایی کنند. کد ژنتیکی که نقشه‌برداری سه‌گانه نوکلئوتیدها به آمینواسیدها می‌باشد، یکی از مهم‌ترین اکتشافات زیستی محسوب می‌شود. بعد از ۳۰ سال بسیاری از عملکردها و پیام‌ها در DNA هنوز ناشناخته مانده است؛ دانشمندان حدس می‌زنند شامل نواحی کدگذاری غیرپروتئینی باشد. پیوند ضعیف DNA فرآیند انعکاس آسان و مولکول مستعد دخالت در سلول می‌کند، بنابراین برای حفظ خطا حالت پایدار، رمزهای دو ـ رشته‌ای (دوتایی) در یک مارپیچ حلزونی هیستون[2] می‌پیچد و می‌توان درک کرد که تکرارهای نوکلئوتیدی مشخص به DNA کمک می‌کند تا در حالت خمیده پیچیده شوند (آب‌گریز بودن بازها) و باعث ایجاد جفت‌هایی پایه‌ای در DNA می‌شود AG، CT همچنین در چرخش هم کمک می‌کنند. یک رشته تلومر با توالی تکراری و رمز TTAGGG که در انتهای کروموزوم قرار دارد، از تخریب اگزونوکلئاز محافظت می‌کند. انتهای کروموزوم توالی مانند TTAGGG داریم، به‌طوری‌که معمولاً داری خاصیت چسبندگی می‌باشند، با حضور تلومر و ساختمان آن چسبندگی از بین می‌رود؛ از این رو تلومر مانند سپری کروموزوم را از ایجاد پیوستگی غیرصحیح و تخریب به وسیله آنزیم‌های اگزونوکلئاز سلول محافظت می‌کند. لذا هر چه پیرتر می‌شویم

[1] Genetic Code
[2] Histon

فعالیت آنزیم تلومراز ضعیف‌تر می‌شود و نمی‌توانند ژن‌ها را در برابر خردشدن محافظت کنند؛ در نتیجه سلول می‌میرد و منجر به پیری می‌گردد. برعکس، وقتی تلومراز بیشتر اظهار می‌شود، تمایل سلول به زندگی و تکرارها بیشتر می‌شود؛ بازها حداکثر تقسیم سلولی را انجام می‌دهند و منجر به سرطان می‌شوند. حالت تناوبی مراحل و تکرارها نقش حیاتی در ثبات ساختار کلی بازی می‌کند. حال با شناخت کوتاه از درون سلول آن را از دیدگاه ریاضی، علی‌الخصوص نظریه کدگذاری بررسی می‌کنیم.

برای تجزیه و تحلیل DNA، نمایش ریاضی نوکلئوتیدها $\{A, T, C, G\}$، اولین گام اساسی می‌باشد. برای آنالیز و بررسی نوکلئوتیدها، ساختار پیریمیدین و پورین را در نظر بگیریم. این موضوع مطرح می‌شود که چرا طبیعت از چنین الفبایی استفاده می‌کند؟

آناستازی[1] نمایش اعداد مختلط را برای نوکلئوتیدها پیشنهاد کرد. البته نمایش مختلفی برای بررسی توالی وجود دارد، که به عنوان مثال می‌توان برای این ۴ پایه، نمایش ساده با عملیات پیمانه‌ای را بیان نمود: $A = ۰, G = ۱, C = ۲, T = ۳$ روش‌های آماری نمادین از مدل‌های مارکو[2] برای نمایش حالت‌های نوکلئوتیدها و پیش‌بینی دنباله ژنی استفاده می‌کند (مرجع [۸] را ببینید).

برای نمایش نوکلئوتیدها لازم است که عملیات ریاضی قطعی روی مجموعه متناهی از عناصر انجام گیرد؛ میدان متناهی بر گروه و حلقه ترجیح داده می‌شود. در واقع یک تناظر بین نوکلئوتیدها و میدان متناهی از مرتبه ۴ در $GF(۴)$ پیشنهاد می‌دهیم. با توجه به خاصیت جالب توسیع میدان‌ها، عناصر را با اعداد صحیح برچسب می‌زنیم و چند جمله‌ای اولیه را روی میدان $GF(۲)$ به فرم $\alpha^۲ + \alpha + ۱ = ۰$ بررسی کنیم.

[1] Anastassiou
[2] Markov

جدول ۱.۲: محاسبات در میدان $GF(۴)$

$+$	۰	۱	۲	۳
۰	۰	۱	۲	۳
۱	۱	۰	۳	۲
۲	۲	۳	۰	۱
۳	۳	۲	۱	۰

$\times$	۰	۱	۲	۳
۰	۰	۰	۰	۰
۱	۰	۱	۲	۳
۲	۰	۲	۳	۱
۳	۰	۳	۱	۲

تفسیر هندسی این نمایش محک علمی بودن تحقیق قلمداد می‌شود.

توجه کنید که داریم

$$\begin{aligned}\frac{Z_۲[x]}{\langle x^۲+x+۱\rangle} &= \Big\{f(x)+\langle x^۲+x+۱\rangle | f(x)\in Z_۲[x]\Big\}\\ &= \Big\{\alpha_۰+\alpha_۱x+\langle x^۲+x+۱\rangle | \alpha_۰,\alpha_۱\in Z_۲[x]\Big\}\\ &= \{۰,۱,x,۱+x\}\end{aligned}$$

$$x^۲+x+۱=۰\to x^۲=x+۱$$

$$x^۳=x.x^۲\to x.(x+۱)=x^۲+x=-۱\equiv ۱$$

$$x^۴=x^۳.x=۱.x=x$$

با توجه به تناظر بین نوکلئوتیدها و عناصر میدان داریم:

$$۰\to ۰\to A$$

$$۱\to ۱\to G$$

$$۲\to x\to C$$

$$۳\to ۱+x\to T.$$

۳.۲ تعیین کد خطی زمینه

با تصحیح خطا فرآیند تکرار، پیوند بین جفت بازها و رشته‌ها فشار کمتری از لحاظ بی‌نظمی نسبت به قبل می‌سازد. کد ژنتیکی سخت‌ترین قواعد را معرفی می‌کند. ۶۴ نوکلئوتیدها با

توجه به الفبای چهارتایی بازها و کدون سه‌تایی (۴ × ۴ × ۴ = ۶۴) انواع آمینواسیدها را به وجود می‌آورند. از ترکیب آمینواسیدها که تاکنون ۲۰ نوع آن‌ها شناخته شده‌اند؛ پروتئین‌ها کدگذاری شده به وجود می‌آید. سوالی که مطرح می‌شود؛ آیا بلوک یا یک کد جهانی در دنباله‌ی دوباره بازخوانی وجود دارد؟

در گذشته مک نودل[1] زوج آزمایی در پیوندهای شیمیایی با ۴ پایه را فرض کرده بود، این موضوع مقدمات را برای پژوهش‌های بعدی در ساختار کدگذاری توالی به وجود آورد. اولین پژوهش‌ها برای کد تصحیح خطا در DNA با استفاده از روش زوج بیت تنها انجام شد (مرجع [۲۷] را ببینید) در حالی که پژوهش‌های بر پایه نظریه کد می‌باشد، نتایج جالب تحقیقات انگیزه پژوهش‌های جدی‌تر در زمینه اصلاح خطا شد. بنابراین ما نیازمند نگرش کلی برای پیدا کردن k زوج‌بیت که جایگزین هر ترتیبی در هر n و k کد بلوکی از یک رشته DNA باشند. ابتدا روشی را معرفی می‌کنیم که در آن از افراز زیر فضایی نظریه کدگذاری بدون اطلاع از n و k استفاده شود (SP)[2]. عموماً تصحیح خطا کانال‌های ارتباطی مد نظر می‌باشد.

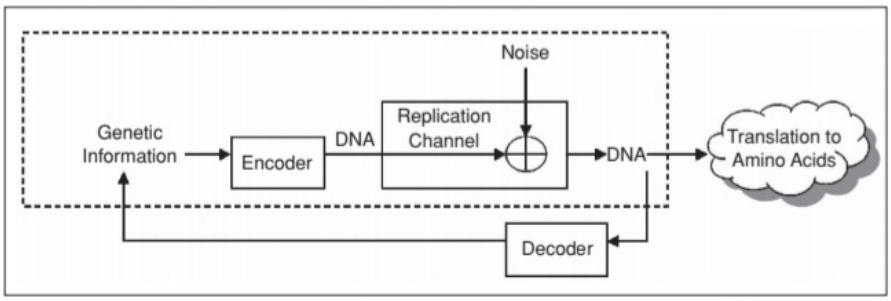
در زیست‌شناسی، جهش‌ها را مورد بررسی قرار می‌دهند و ذکر علت می‌کنند که چنین مبحثی درکدگذاری و مخابرات مورد بحث نیست. قالب نمادی میدان گالوا با ۴ پایه متفاوت استفاده می‌شود. در واقع روش‌های تصحیح خطا در نظریه کدگذاری را برای زیست‌شناسی اعمال می‌کنند.

۴.۲ مدل‌سازی کانال‌های تکراری

امروزه رسانه‌های اطلاعاتی اعم از سیستم‌های اطلاعاتی و دستگاه‌های ذخیره‌سازی داده‌ها به دلیل وجود پارازیت‌ها و دیگر اشکال مداخله به طور کامل قابل اطمینال و اعتماد نمی‌باشد و پارازیت‌ها می‌توانند پیام‌های مخابره شده را مختل کنند. با کدگذاری می‌توان به کشف خطا و تصحیح خطا به صورت بهینه دست یافت. کدگذاری به صورت کدگذاری منبع[3] تغییراتی در منبع پیام داده تا کدهای مناسب جهت انتقال پیام فراهم شود و کدگذاری کانال[4] کد کردن مجدد کلمات کدگذاری منبع قبل از ورود به کانال پارازیت‌دار می‌باشد که به جهت بالا بردن

[1] Mac Donaill
[2] Subspace par
[3] Encoding of source
[4] Encoding of channel

قابلیت تشخیص و تصحیح خطای اعمال شده در کانال پارازیت‌دار صورت می‌پذیرد. از دیدگاه دیگر می‌توان مدل‌های کانال‌های ارتباطی را با روند تکامل DNA تطابق داد. فرض کنید یک ساختار کدگذاری دنباله‌ای وجود دارد که توالی را در مقابل پارازیت‌های که در حین تکرار DNA رخ می‌دهد، محافظت می‌کند.



شکل ۱.۲: مدل‌سازی کدگذاری و کدگشایی DNA

در شکل ۱.۲ DNA وارد کانال تکرار پارازیت‌دار می‌شود؛ این فرآیند در هسته سلول اتفاق می‌افتد؛ وقتی DNA وارد سیتوپلاسم می‌شود، فرآیند ترجمه و تبدیل آمینواسیدها به پروتئین رخ می‌دهد. در مقاله [۴۴] هدف اصلی شناسایی و رده‌بندی هر رشته که احتمالاً در نواحی دنباله‌ها ظاهر می‌شود و شناسایی تمام محدودیت خطی که ممکن در رشته‌ها نمایان شود.

تعریف ۱.۴.۲. فرض کنید S یک زیر مجموعه‌ای از فضای برداری F_q^n باشد مکمل متعامد[1] مجموعه‌ی S که با نماد $S^\perp$ نمایش داده می‌شود. به صورت زیر تعریف می‌کنیم

$$S^\perp = \{V \in F_q^n | \quad V.s = 0; \forall s \in S\}$$

اگر $S = \varnothing$، آن‌گاه داریم: $F_q^n = S^\perp$.

در این بخش روش‌هایی را معرفی می‌کنیم تا نواحی متناوب در سرتاسر فریم برای تسهیل چک کردن به صورت بصری برجسته شود؛ با استفاده روش SP و LD[2] توالی تکراری را بررسی می‌کنیم.

کدواژه از طول n را در نظر می‌گیریم. ماتریس $V = [v_1 v_2 \dots v_N]$ داریم، جایی که v_i ستون‌های از طول n باشد. ابتدا توالی DNA را در ماتریس قرار می‌دهیم؛ پایه مکمل متعامد تولید می‌شود.

[1] Orthogonal complement
[2] Linear Dependance Test

از طرفی در فرایند ترجمه پروتئین‌ها، کدها سه‌تایی خوانده می‌شود و هر کدون یک اسیدآمینه خاص تولید می‌کند. اگر نوکلئوتیدی حذف یا اضافه شود؛ مضربی از سه نباشد، از آن نقطه به بعد قالب خواندن open reading fram می‌شود. هر توالی mRNA سه فریم دارد که تنها یکی صحیح بقیه زود به کدون پایان می‌رسد و تغییر خوانش داده و شیفت پیدا می‌کند.

شکل ۲.۲: جهش درون قالبی نوکلئوتیدها

هم‌ترازی و میزان تطبیق قالب‌ها نسبت به شروع اولیه میزان انحراف در نظر گرفته می‌شود. از روش گرام‌ـاشمیت روی میدان‌های متناهی استفاده می‌شود و دنباله‌ای از بردارها بدست می‌آید. در واقع روشی را پیشنهاد می‌دهیم تا پایه مکمل متعامد برای منطقه محلی از داده‌ها تولید کند $t_i = Gv_i$ جایی‌که

$$G = \begin{pmatrix} e_۱ \\ e_۲ \\ \vdots \\ en \end{pmatrix}.$$

در این روش توالی DNA به صورت ستونی در ماتریس قرار می‌گیرد. در واقع یک تبدیل مختصاتی داریم که t_i ستون دارد و پایه جدید را در v_i قرار می‌دهیم، مشروط به این که الگوریتم اجرا شود الگوریتم گرام ـ اشمیت یکسری نواحی زائد مربوط به کدخطی تشخیص می‌دهد، مولفه‌های صفر از طول n داریم که با عملیات سطری‌پلکانی از صفر (زائد) صرف نظر می‌شود.

۵.۲ الگوریتم رده‌بندی رشته های DNA

۱. گرام–اشمیت (روش ۲.۱ ببینید) روی v_i اعمال و پایه متعامد $\{e_1, e_2, \ldots, e_n\}$ حاصل می‌شود، جایی‌که $j \geq n$. ماتریس تبدیل G بدست می‌آید.

۲. رشته‌های مؤلفه پایه $\{t_1 \ldots t_{i-j}\}$ را برای تمام فرم قالب ممکن تجزیه می‌شود.

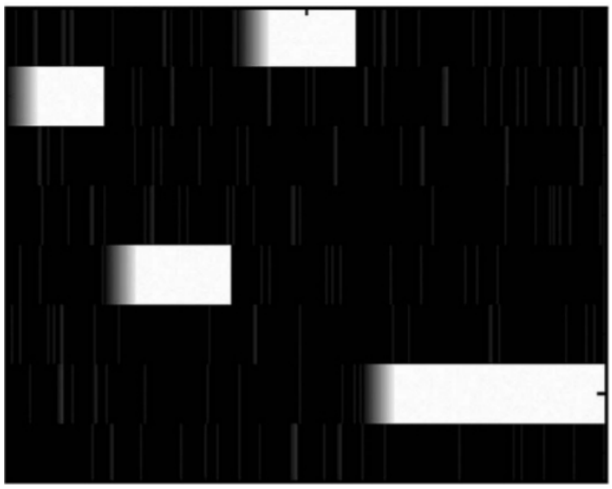
۳. صفرها t_i توجه کنید که ناحیه دارای افراز زیرفضایی یکسان و زائد نشان می‌دهند. میزان اطمینان احتمال مجموعه‌های بردارهای تصادفی که دارای افراز مشابه‌ای از دنباله‌ها می‌باشند.

روش SP قادر به کشف رده‌بندی افرازهای زیرفضای خطی هر رشته می‌باشد، مشروط به این که الگوریتم برای قالب خواندن هر طول کدواژه‌ای که در دسترس اجرا شود.
به‌عنوان مثال در شکل ۳.۲ جهش درون قالبی را برای پرسلولی و در شکل ۴.۲ برای باکتری اشرشیاکلی[1] مشاهده می‌کنید. رشته آزمایش برای اشغال کردن یک ۵ بعدی از فضای ۸ که یک $(8,5)$ کدبلوکی در $GF(4)$ می‌باشد با اجرای الگوریتم روی رشته ۸ و نمایش تصویر محرمانه سرتاسر رشته، توانمندی الگوریتم در قالب خواندن نشان می‌دهد. هرگاه جهش قالب خواندن رخ دهد، محل و منطقه زیرفضایی به راحتی تغییر و به سطر متناظر در دیگرام می‌رود. اجرای الگوریتم در شکل ۳.۲ مشاهده کنید.

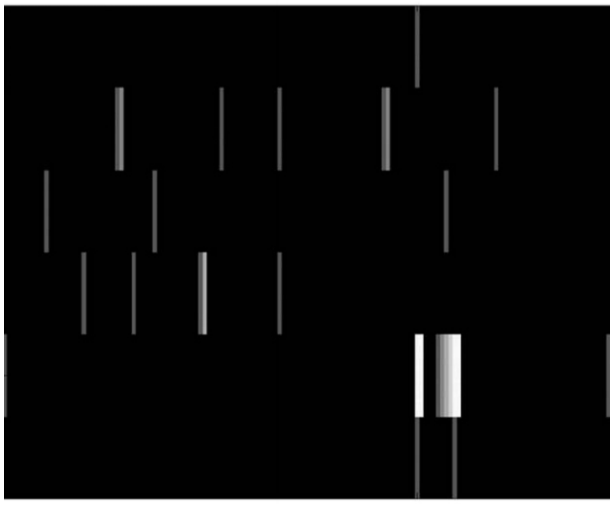
یک افراز زیرفضای خطی برای $8,5$ بلوکی در مناطق سفید داده‌ها نمایان می‌شود. در روش SP باید توجه داشت افراز زیرفضای به دستگاه صحیح جهت‌دهی شده باشد؛ در غیر این صورت پایه تغییر می‌کند و نمی‌تواند کد بلوکی یکسان داشت. فرض کنید جهت دهی دستگاه صحیح باشد؛ کد بلوکی ناحیه‌ای و منطقه‌ای کشف می‌شود ولی کد کلی نداریم. در شکل ۴.۲ باکتری $E.coli$ کدهای محلی و کوچک را به صورت نوارهای روشن داریم که در شکل ۳.۲ به صورت بلوکی می‌باشد.

می‌توان نتیجه گرفت یک کد بلوکی خطی پایدار بدست نمی آید که بتوانیم برای تمام رشته ها کارا باشد اما یکسری نواحی با زیرفضایی یکسان برای قالب‌های پی‌درپی، با توجه به شدت

[1] Escherichia coli



شکل ۳.۲: تست SP بر $(۸,۵)$ کد بلوکی خطی در $GF(۴)$



شکل ۴.۲: تست SP بر روی باکتری $E.coli$

و چگالی روشن دیده می‌شود. در الگوریتم SP لازم است مختصات G صحیح جهت‌دهی شده باشد تا کد بلوکی را نمایش دهد مشروط به این‌که ساختار قالب از ابتدا مشخص باشد. روش SP ساختار کدگذاری بلوکی نشان می‌دهد، در مقابل جهش‌ها پایدار نمی‌باشد. الگوریتم مشروط به جهت‌دهی صحیح می‌تواند توالی تکرار را نشان دهد [۴۳].

۶.۲ کشف تکرار و نواحی زائد DNA

با مطالعه ساختار DNA به توالی تکراری آن پی می‌بریم. در روش LD از روش‌های ترکیبیاتی و پیچیده استفاده می‌شود تا توالی تکراری که ترتیب خاص ندارند پیدا شود. بعلاوه از قالب میدان متناهی، ماتریس، وابسته خطی بودن که ابزار قدرتمند جبر برای مدل‌سازی توالی تکراری استفاده می‌شود. می‌توان گفت در برابر انواع جهش‌ها ۱۷.۴.۱ رفتار الگوریتم LD چطور خواهد بود. الگوریتم LD یکی از انعطاف‌پذیرترین الگوریتم‌ها با آستانه تشخیص مختلف و توانایی نمایش کارامد موقعیت نوکلئوتیدها می‌باشد. روش‌های جبری خطی ذکر شده در روش SP برای شناسایی توالی تکرار در روش LD گسترش می‌دهیم. از نظر بیولوژیکی، آزمون SP برای ساختار دقیق برنامه‌نویسی بلوک می‌باشد، در حالی که آزمون LD می‌تواند افزونگی سخت مانند تکرارهای تقریبی را نیز تشخیص دهد. آزمون LD مناطق زائد محلی را تعیین می‌کند و نقطه شروع خوبی برای تجزیه و تحلیل بیشتر مانند تکرارهای پشت سر هم می‌باشد.

۷.۲ تست LD

در روش LD [۴۳] یک پنجره با طول N^2 از داده‌ها به صورت ماتریس $N \times N$ تغییر شکل می‌یابد. این ماتریس حداکثر N بعد را اشغال می‌کند. در رتبه LD از هر پنجره $N \times N$ محاسبه رتبه براساس الگوریتم حذف گوسی[1] می‌باشد. پس داده‌ها با هر بار تکرار قالب N تا طول افزایش می‌یابد و درنتیجه یک پنجره $N \times N$ به آرامی در حال حرکت ایجاد می‌شود که هر بار توسط N نوکلئوتید طی می‌شود. طرح کلی روش به صورت زیر است:

الف) برای تجزیه و تحلیل قالب از طول N و جمع‌آوری N بردارهای متوالی برای تشکیل پنجره $N \times N$

ب) محاسبه رتبه ماتریس $N \times N$.

ج) اضافه کردن یک قالب برای هر تکرار.

[1] Gaussian-elimination

د) توجه داشته باشید کمبود رتبه با افزایش I سازگار که خود این روش یک وابستگی خطی منطقه‌ای است و تفاوت در اشغال بعد بین پنجره همپوشانی پیدا می‌کند.

در شکل ۵.۲ تست LD برای $N = ۱۳۵$ یست کروموزوم[1] نشان می‌دهد. متناسب با طول و سطح کمبود رتبه پنجره $N \times N$ متوالی که هر کدام از یک قالب خاص شروع می‌شود. تعداد دو منطقه از ژن $FLO۹$ نشان می‌دهد که با روش LD بسیار تکرار می‌شود. ما روشی را معرفی می‌کنیم که نواحی متناوب در سرتاسر فرم قالب برجسته شود . در شکل ۵.۲ محور X که بیانگر فریم و محور Y افست‌ها را نشان می‌دهد . تمام شیفت قالبی آزمایش شده است؛ با بررسی جبری این رابطه برقرار است اگر زیرفضا از رتبه $N - ۱$ پیدا شود، در قسمت خاکستری تیره نمایش داده شود. وقتی رتبه زیرفضایی را کاهش می‌دهد وابسته خطی را نمایشگر I شکل سایه روشن‌تر می‌شود. در واقع کاهش و افزایش میزان روشنایی گراف تابعی از دو متغییر استحکام و طول ناحیه زائد است ابتدا الگوریتم روی ناحیه زائد هفده هزار برای $N = ۱۳۵$ بیان می‌کنیم و تعداد پایه‌هایی که ناحیه‌های ردیف تولید کند می‌تواند شماره شکل $(N \times N) + N \times$ [۴۳]. ماتریس $N^۲$ اگرچه نواحی دارای نقص رتبه‌ای حداکثر یک یا دو واحد می‌باشد تا فول رنک[2] شوند؛ نقص رتبه‌ای می‌باشد، ولی خیلی وابسته خطی نیست و قالب‌های نوکلئوتیدهای تقریبا یکسان و مشابه حاصل می‌شود. [۴۳].

در شکل ۵.۲ و ۶.۲ تکرار پایه را برای N متفاوت بررسی شده است؛ با استفاده از روش LD توالی تکراری و زائد را نشان می‌دهد. توالی با $N = ۱۳۵$ مربوط مخمر یست کروموزم[3] می‌باشد (شکل ۵.۲) . دنباله ژنوم انسانی و تکرارهای پشت سر هم برای $N = ۱۹$ بررسی می‌شود. در شکل ۶.۲ آزمون LD می تواند ناحیه‌ای که طولانی‌ترین تکرار را دارد با آزمایش تمام قالب خوانش نشان دهد؛ ناحیه‌ی زائد را به‌صورت بیرون زدگی مشاهده می‌کنید. (شکل ۶.۲ را ببینید) الگوریتم در مقابل جهش‌ها پایدار می‌باشد و می‌تواند تکرارهای غیرمنتظره را پیدا کند و مناطق زائد را به صورت بیرون زدگی نشان دهد [۴۳] .

عوامل محیطی، دما، تابش الکترومغناطیسی در میزان خطا و جهش تأثیر می‌گذارد.

[1] Yeast Chromosome
[2] Full rank
[3] Yeast chromosome

شکل ۵.۲: تست LD یست کروموزوم

شکل ۶.۲: تست LD ژنوم انسان

علاوه برجهش‌ها که توسط مولکول بدترکیب در DNA ایجاد می‌شود، تکرار رفت و برگشت می‌تواند خطاهایی را معرفی کند که در ساخت یاخته پدیدار می‌شود (میکروساتلیت). در واقع زمانی که DNA کپی می‌شود مکانیسم تصحیح خطا دارد؛ ساختار حلزونی جدا می‌شود و به دو رشته مجزا منشعب می‌شود و بازها به هم می‌چسبند تا پله‌های جدید را کامل کنند. وقتی یک جهش جانشینی که معمولا یک پورین با یک پیریمیدین جایگزین، و یک پیریمیدین می‌شود و این عدم تطابق یک گره در گسترش ایجاد می‌کند.

در بدن انسان جهش وجود دارد و رفتار الگوریتم LD در برابر انواع جهش‌ها بررسی می‌شود. نتایج پژوهش هوث[1] برای $N = ۴۸$ در شکل ۷.۲ می‌بینید.

حروف بزرگ نشان دهنده بخش‌های محافظت شده است. حروفی که زیر آن‌ها خط کشیده شده است از قالب قبلی اضافه شدند. حروفی پررنگ نشان دهنده نواحی که بعد از حذف که در فرم قبلی رخ داد، حفظ می‌شود. حروف کج و ایتالیک نواحی قبل حذف هستند. حرف کوچک، نمایش داده خطا جایگزین است. حروف کم رنگ بخش‌هایی که یک جایگزین پایه

[1] Hauth

شکل ۷.۲: الگوریتم LD برای $N = ۴۸$

چندگانه برای یک پایه مشخص رخ داده است. الگوریتم در فرم قالب متفاوت است؛ چون به صورت تقریبی می‌باشد و فرم خوانش اگر مضرب N نباشد جهش به وجود می‌آید، این در حالی؛ توالی در DNA دقیق است. الگوریتم تقریبا توالی را بررسی می‌کند؛ لذا می‌تواند نواحی نزدیک را شناسایی کنند ولی دقیق نیست. حروف کوچک و خاکستری روشن نواحی را نشان می‌دهد که در آن نوکلئوتیدها به نوکلئوتیدهای دیگر تبدیل می‌شود و جهش جایگزینی رخ می‌دهد. پایه خاکستری روشن اضافه شدن مانند دی نوکلئوتیدها را نشان می‌دهد.

برتری روش LD در این است؛ علارغم خطاها که در توالی رخ می‌دهد؛ نسبت به الگوریتم‌های دیگر در برابر جهش‌ها پایدارتر است و گمراه نمی‌شود. می‌تواند توالی مانند مضرب‌های توالی $N = ۲۴$پیدا کند. فرض کنید یک تصحیح خطا در ناحیه کدینگ و غیر کدینگ وجود دارد به برهان خلف فرض می‌کنیم، نرخ جهش از یک ژنوم به ژنوم دیگه تغییر کرده باشد در این صورت برای آن واحد، رفتار خاص آن ناحیه در نظر گرفته می‌شود.

هم اکنون از دیدگاه دیگر کد را بررسی کرد. دو راه برای تعریف کد وجود دارد:

۱. مشخص کردن قوانین ساختن آن، همانطوری که مهندسی علم ارتباطات و اطلاعات معمولاً از این استفاده می‌کند.

۲. مشخص کردن محدودیت مورد نیاز برای کدواژه‌ها.

نگرش دوم مناسب‌تر برای پدیده‌های طبیعی می‌باشد. در حالت‌هایی کد نرم شبیه تعریف یک ماتریس کنترل توازن به جای ماتریس مولد در نظریه کدگذاری می‌باشد.

به عنوان مثال؛ در کدهای همینگ[1] با ماتریس کنترل توازن کار می‌کنیم. برای درک بهتر کدهای نرم آن را با یک زبان طبیعی مقایسه می‌کنیم. در زبان طبیعی محدودیت‌های بسیار از نظر آواشناسی و معنایی داریم. همچنین محدودیت‌هایی که از نظر قواعد گرامری و صرف و نحو در زبان وجود دارد. زبان‌های طبیعی دارای توانایی بالایی از لحاظ تصحیح خطا از هر دو جنبه شفاهی و نوشتن که به صورت سطوح سلسله مراتب مختلف عمل می‌کند در هر سطح محدودیت پی‌درپی می‌باشد. برای مثال، محدودیت آوایی خیلی اساسی و غیر قابل انعطاف برای یک زبان که عمدتاً براساس عرف اجتماعی و عموم می‌باشد.

۸.۲ شبکه‌های تنظیم ژن

در این بخش ژنوم و کدها در حالت خاص بررسی می‌شود. از کدهای با ماتریس توازن کم چگال (کدهای LDPC[2]) استفاده می‌کنیم.

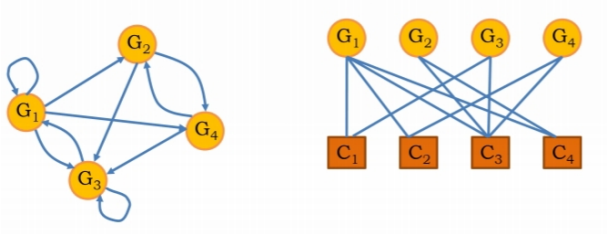
GRNs ۱۶.۴.۱ ژن ها مستقل از هم نیستند. بیان یک ژن بر ژن دیگر تاثیر می‌گذارد و اصطلاحاً ژن بیان[3] شده است. زمانی که یک ژن بیان می‌شود، بر بیان یا سرکوب ژن دیگر تاثیر می‌گذارد، با تولید آنزیم‌ها و پروتئین هایی که خودشان را به تنظیم کننده ژن می‌چسباند از این رو بیان یک ژن منتهی به یکسری واکنش‌هایی در سلول می‌شود.

برای بررسی شبکه ژن، یک گراف ژنتیکی را در نظر می‌گیریم. در گراف N ژن به عنوان N تا راس می‌باشد و بیان ژن‌ها و تاثیر روی همدیگر را یال‌های جهت‌دار (کمان) نشان می‌دهیم.

بیان G_i بر بیان G_j تاثیر می‌گذارد. بیان G_i را به صورت تابع f_i متناظر می‌کنیم.

در شکل ۸.۲ f_i مجموعه از ژن‌ها می‌باشد، $(f_i = ۰, ۱)$. اگر $f_i = ۱$ ژن بیان و اگر $f_i = ۰$ ژن سرکوب شده است. G_i معمولا در گراف‌های ژنتیک درجات ورودی رئوس نسبتاً کوچک و کم چگال می‌باشد. می‌توان گفت گراف‌های ژنتیکی خیلی یال ندارند به همین علت کدی که برای آن در نظر می‌گیریم کم چگال و کم یال می‌باشد. به علاوه گاهی اوقات در گراف

[1] Hamming Coding
[2] Low Density Parity Check
[3] Expressed



شکل ۸.۲: گراف ژنتیکی

یکسری یال‌های طولانی نیز وجود دارد که ژن‌هایی که فاصله بیشتری از هم دارند و بر هم تاثیر می‌گذارند. هدف تجزیه و تحلیل دقیق‌تر گراف ژنتیک به یک گراف دوبخشی می‌باشد.

شبکه‌های تنظیم ژن: فرض کنید شبکه تنظیم ژن هیچ اختلالی نداشته باشد، حالت ورودی X داریم، فرض X همگرایی به Y دارد. یک آشفتگی d، که متناظر با خطاها X می‌باشد. علاقمندیم در شبکه‌هایی با آشفتگی کوچک و نقطه جذبی صحیح Y؛ مراحل تکامل شبکه تنظیم ژن و بیان و سرکوب در سیتوپلاسم بررسی کنیم. یک ماتریس $N \times N$ با درایه‌های i و j؛ داریم:

$$\frac{\partial f_i}{\partial x_j} = f(x_1, \cdots, x_j = 0, \cdots, x_N) \oplus f(x_1, \cdots, x_j = 1, \cdots, x_N)$$

$$F = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \cdots & \frac{\partial f_1}{\partial x_N} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \cdots & \frac{\partial f_2}{\partial x_N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_N}{\partial x_N} & \frac{\partial f_N}{\partial x_N} & \cdots & \cdots & \frac{\partial f_N}{\partial x_N} \end{bmatrix}$$

$$F(t) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(t) & \frac{\partial f_1}{\partial x_2}(t) & \cdots & \cdots & \frac{\partial f_1}{\partial x_N}(t) \\ \frac{\partial f_2}{\partial x_1}(t) & \frac{\partial f_2}{\partial x_2}(t) & \cdots & \cdots & \frac{\partial f_2}{\partial x_N}(t) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_N}{\partial x_N}(t) & \frac{\partial f_N}{\partial x_N}(t) & \cdots & \cdots & \frac{\partial f_N}{\partial x_N}(t) \end{bmatrix}$$

$$d(t+1) = F(t) \otimes d(t)$$

$$d(t) = F(t-1) \otimes d(t-1)$$

$$F(t-۱) \otimes F(t-۲) \otimes d(t-۲) \cdots$$

$$F(t-۱) \otimes F(t-۲) \cdots F(۱) \otimes d(۰)$$

فرض کنید خطا در کدگشایی ژن G_i رخ دهد. این اتفاق یک آشفتگی در شبکه به وجود می‌آورد. اگر شبکه داری قابلیت تشخیص باشد این آشفتگی را مغلوب می‌کند و باعث پیشگیری بسیاری از بیماری‌های صعب‌العلاج می‌شود. دانشمندان حدس می‌زنند محل اتصال گره ژنتیکی ساماندهی شده است و فقط ژن‌های معتبر کدواژه می‌شوند. برخی شواهد بیانگر حضور نوکلئاز در مسیر پر نویز تکثیر و همچنین نقش تصحیح هیستون‌ها، می‌توان گفت سوء عمل مکانیزم دوباره خوانی[۱]؛ می‌تواند علت بسیاری از بیماری‌ها باشد.

۹.۲ کدگذاری و کدگشایی عصبی

هدف از این بخش، مطالعه کدگذاری عصبی [۲] با مدل کردن سیستم‌های عصبی به‌عنوان کانال‌های ارتباطی است. کدگذاری نورونی یکی از مباحث مرتبط به علوم اعصاب و مغز است که به توصیف ارتباط بین محرک و پاسخ گروه نورون‌ها و فعالیت الکتریکی اعصاب می‌پردازد. بر اساس این نظریه که اطلاعات حسی در مغز توسط شبکه‌ای از نورون‌ها بیان می‌شود، دانشمندان حدس می‌زنند که سلول‌های عصبی می‌توانند اطلاعات دیجیتال و آنالوگ را کد کنند. کدگذاری یک محتوای اطلاعاتی، به معنی تمایز و تفکیک یک حالت از محتوای اطلاعاتی کدگذاری، از حالت‌های ممکن دیگر برای آن محتوای اطلاعاتی می‌باشد. در واقع پاسخ نورون‌ها، فعالیت الکتریکی آن‌ها در نظر گرفته می‌شود؛ که شلیک نورونی یا قطارها اسپایک[۳] (پتانسیل عمل نورون) می باشد. یک قطار شلیک توسط دنباله‌ای از اعداد که معرف زمان‌ها رسیدن پتانسیل عمل به نقطه مورد نظر است بررسی می‌شود. محتوای اطلاعاتی، تمایز و تفکیک نوع محرک است که بر اساس خوانش فعالیت الکتریکی آن گروه از نورون‌ها انجام می‌شود. این عمل خوانش را می‌توان کدگذاری و سپس کدگشایی در نظر گرفت، که در ابتدا و انتها یک کانال ارتباطی انجام می‌شود. ریشه‌های ایده دیدگاه کانال ارتباطی از رویکرد نظریه اطلاعات نشات گرفته است.

[1] Proof Reading
[2] Neural Coding
[3] Spike Trains

اگر چه بعضی شواهد به نفع وجود روش‌های کنترل خطا در DNA وجود دارد، اما بحث های مخالف این مکانیسم هم مطرح می شود. هیچ کدام از الگوریتم و مدل های قبل گواهی بر اصلاح خطا در DNA نیست و فقط پیشنهاد می‌شود که چنین مکانیسم باید وجود داشته باشد. علاوه بر این برخی آزمایش‌ها در یافتن کدگذاری سلول ناموفق بوده است. اکنون نوعی مدل‌سازی عصبی شبیه آن چه در نظریه اطلاعات می‌باشد؛ بیان می‌کنیم.

ابتدا یک ورودی با احتمال توزیع $p(x)$ در نظر می‌گیریم. ورودی کانال با توزیع احتمال $p(y|x)$ فرستاده می‌شود منجر به خروجی y با توزیع احتمال $p(y)$. در سیستم‌های عصبی ورودی محرک و کانال ارتباطی اعصاب و خروجی پتانسیل عمل است. با استفاده از نظریه کدگذاری توالی $Y^{۱}$ و توالی پراکنده X ، یک دسته عمومی طولانی تعیین می‌کنند. ورودی و خروجی رده های یکسان مشترک شکل می‌دهند. حال برای شناسایی رده‌های یکسان، باید I، انتروپی H بین X و Y پیدا کنیم. یک چالش بزرگ حجم بالا از داده‌های مورد نیاز است. برای غلبه بر این مشکل بین توزیع احتمال ممکن آن که بیشترین اطلاعات بین محرک و پاسخ محافظت می‌کند انتخاب می‌شود. یک کوانتیزه تصادفی برای ترسیم توالی خروجی Y با نمونه جدید کوانتیزه Y_N با سطوح سنجش N طبق توزیع احتمال q با توجه به توزیع تعیین شده $q(Y_N|Y)$ داریم. نرخ کد معادل یک کوانتیزه قطعی است. کیفیت اندازه گیری با رابطه زیر بدست می‌آید:

$$D_I(Y;Y_N) = I(X;Y) - I(X;Y_N) \qquad (۲)$$

اکنون مسئله کوچک کردن D_I است که معادل زیر می‌باشد:

$$\max_{q(Y_N|Y)} H(Y_N|Y) \qquad (۳)$$

مشروط به این‌که :

$$Deff = I(X;Y_N) \geq I_{\circ} \qquad (۴a)$$

$$\sum_{\nu=۱}^{N} q(y_\nu|y) = ۱; \qquad \forall y \qquad (۴b)$$

جایی‌که $I_{\circ}$ یک آستانه دقت کوانتیزه برای مثال $I(X;Y_N) = I_{\circ}$؛ بنابراین برای یافتن ۳ نیازمند است کوانتیزه که حداقل اطلاعات $I_{\circ}$ را درباره X حمل می‌کند، پیدا کنیم. با اجرا داده $I_{\circ}$ بیشتر به نقطه $I_{\circ}^{max}$ دست می‌یابیم. مؤلفان از ضریب لاگرانژ استفاده می‌کنند تا ۳ را

حل کنند. به خصوص شرایط کاروش_ کان_ تاکر KKT برای تنظیم مشتق‌های تابع هدف لاگرانژ تا به صفر برساند، شکل بسته $q(Y_N|Y)$ در معادله ۵ داده می‌شود.

$$q(y^\nu N|y_k) = \frac{e^{-\beta \frac{\nabla D_I^{\nu^k}}{p(y_k)}}}{\sum_{\nu=1}^{N} e^{-\beta \frac{\nabla D_I^{\nu^k}}{p(y_k)}}} \qquad (۵)$$

در جایی که β ضریب لاگرانژ برای ۴a و $q(y^\nu N|y_k)$ احتمال y_k متعلق به رده یکسان ν و $\nabla D_I^{\nu^k} = \partial D_I / \partial_q(y_N^\nu|y_k)$. D_I مقدار N می‌دهد. با داشتن N ساختن $q(y^\nu N|y_k)$ شروع می‌کنیم. بنابراین یک توزیع مشابه $q(y_N^\nu|x)$ می‌توانیم بسازیم که احتمالی شبیه توالی ورودی دارد. اگر D_I کوچک، پاسخ یکسان با رده $y^\nu N$ برابر

$$y^\nu = \{y_k|q(y_N^\nu|y_k) \simeq ۱, \forall k\} \qquad (۶)$$

در این صورت اگر ما همه x را گروه‌بندی کنیم خروجی این رده:

$$X^\nu = \{x|q(y^\nu N|x) \simeq ۱\} \qquad (۷)$$

بنایراین ما توزیع احتمال $q(y^\nu|x^j)$، که احتمالی یک عضو از رده یکسان j^th ترسیم شلیک نورونی یا قطارها اسپایک ν از رده یکسان ν. مدل اشاره شده توانایی توزیع رده‌های یکسان را دارد که در آنالیز عصبی و پاسخ های محرک بسیار کمک کننده است.

همچنین می توان اشاره کرد، به مک دونیل [1]، که یک کد زوج‌آزمایی را که از ترکیب نوکئوتیدها کشف کرد. در این مدل نوکلئوتیدها یک کد زوج آزمایی ۴ تایی را تشکیل می‌دهند [۳۱]. همچنین نویسندگان مقاومت متوسط هر کد در برابر خطاهای تصادفی بین کدون‌ها بررسی کردند و می‌توان گفت فقط یکی از کدها از حجم میلیون کد تصادفی از کد ژنتیک استاندارد بهتر عمل کرده است و احتمالا واگذاری کدون‌ها در مسیری که خطاهای تفسیر را کم می‌کنند، بهینه می‌شود.

[1] MacDonaill

فصل ۳

کدهای DNA برگشت‌پذیر

۱.۳ مقدمه

مقالات و مطالعات متعددی در مورد کدهای DNA و همچنین محاسبات ساختار DNA وجود دارد، اما از آن جایی‌که در این فصل به طور خاص به بررسی کدهای دوری اریب و کدهای DNA متمرکز شده ایم، لذا لازم است با بیان برخی مفاهیم و تعاریف، ساختار کدهای دوری، کدهای دوری اریب و کدهای DNA برگشت‌پذیر را بررسی کنیم . همچنین در مورد ساختار برگشت‌پذیر DNA به وسیله‌ی کدهای دوری اریب در میدان $\mathbb{F}_{۴^{۲s}}$ بحث می‌کنیم. ملاحظه می‌کنیم چطور وجود و خواص کدهای دوری اریب به ساختار کدهای DNA برگشت‌پذیر ارتباط پیدا می‌کند.
$[n,k]$ کد دوری روی $GF(q)$ را می‌توان به عنوان یک ایده‌آل روی گروه جبری $GF(q)$ در نظر گرفت، که $C = \langle g \rangle$ گروه دوری از مرتبه n با مولد g است. کدهای دوری رده مهمی از کدها از دیدگاه نظری و عملی هستند. کلید توصیف روش ساخت کدهای دوری روی حلقه R این است که کدهای دوری را همانند ایده‌آل درحلقه چندجمله‌ای $\frac{R[X]}{\langle x^n-۱ \rangle}$ ببینیم که n طول کد است. به این منظور بدست آوردن مقسوم علیه $x^n - ۱$ موثرخواهد بود، اما زمانی که مشخصه حلقه نسبت به طول کد اول است ریشه مکرر حاصل می‌شود، بنابراین $x^n - ۱$ عامل منحصر به فرد روی حلقه ندارد. برای کدها روی $\mathbb{Z}_۴$ این مورد متناظر موردی است که طول زوج است.

تعریف ۱.۱.۳. کد C را یک کد دوری[۲] نامیم هرگاه

الف) خطی باشد

[2] Cyclic Codes

ب) هر انتقال دوری از یک کدواژه‌ی C، باز هم کدواژه ای از خود C باشد، به عبارت دیگر، اگر $c = (a_1, \ldots, a_n) \in C$ آن‌گاه $\acute{c} = (a_n, a_1, a_2, \ldots, a_{n-1}) \in C$.

مثال ۱.۱.۳. کد $C = \{000, 101, 011, 110\}$ یک کد دوری دودویی می‌باشد.

تعریف ۲.۱.۳. فرض کنید $h(x) = h_0 + h_1 x + \cdots + h_k x^k$ چندجمله‌ای آزمون باشد، آن‌گاه چندجمله‌ای $\bar{h}(x)$ را به صورت زیر تعریف می‌کنیم:

$$\bar{h}(x) = h_k + h_{k-1} + \cdots + h_1 x_{k-1} + h_0 x^k$$

$\bar{h}(x)$ را چندجمله‌ای وارون نامیم.

قضیه ۱.۱.۳. فرض کنید C یک کد دوری روی R_n باشد؛

۱. یک چندجمله‌ای تکین منحصر به فرد $g(x)$ از کوچکترین درجه در C وجود دارد؛

۲. $C = \langle g(x) \rangle$

فرض کنید $g(x)$ و $h(x)$ دوچندجمله‌ای تکین از کم‌ترین درجه m باشند. چون C دوری است و نسبت به جمع بسته می‌باشد و از طرفی تفاضل نیز قرینه جمع است. $r(x).h(x) - g(x) \in C$ چندجمله‌ای است متعلق C که درجه‌اش از m کم‌تر و این تناقض است.

هر $a(x) \in C$ را می‌توان به صورت مضربی از $g(x)$ نمایش داد. یعنی $a(x)$ بر $g(x)$ بخش‌پذیر است.

قضیه ۲.۱.۳. فرض کنید C یک کد دوری با چندجمله‌ای مولد $g(x) = g_0 + g_1 x + \cdots + g_r x^r$ باشد؛ $\dim C = n - r$ و ماتریس مولد آن به صورت زیر است:

$$G = \begin{bmatrix} g_0 & g_1 & \cdots g_r & 0 & 0 & \ldots 0 \\ 0 & g_0 & g_1 & \cdots g_r & 0 & \ldots 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots g_0 & g_1 & \cdots g_r \end{bmatrix}$$

برهان. ماتریس مولد کد C می‌نویسیم. در سطر اول چند جمله‌ای $g(x)$ قرار می‌دهیم که $r+1$ درایه 0 و 1 حاصل می‌شود. که r بزرگ‌ترین توان است و در قضیه ۱.۱.۳ ثابت کردیم که $g_0 = 1$ است. اگر $r < n$ باشد، یعنی اگر درجه چندجمله‌ای مولد کم‌تر از طول باشد برای دست یافتن به طول n به تعداد $n-(r+1)$ صفر پس از نوشتن درایه‌های چندجمله‌ای مولد $g(x)$ قرار می‌دهیم. در نوشتن سطرهای دیگر ماتریس مولد، سطر اول را $n-(r+1)$ بار انتقال داده‌ایم و خود سطر اول نیز موجود است. پس بعد کد $n-(r+1)+1 = n-r$ خواهد بود. □

قضیه ۳.۱.۳. فرض کنید C یک $[n,k]$-کد دوری که چندجمله‌ای آزمون آن به صورت زیر $h(x) = h_0 + h_1 x + \cdots + h_k x^k$ باشد آن‌گاه

$$H = \begin{bmatrix} h_k & h_{k-1} & \cdots h_1 & h_0 & 0 & 0 & \cdots 0 \\ 0 & h_k & h_{k-1} & \cdots h_1 & h_0 & 0 & \cdots 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots 0 & h_k & h_{k-1} & \cdots h_1 & h_0 \end{bmatrix}$$

H ماتریس مولد $C^\perp$ می‌باشد و $C^\perp$ کد تولید شده توسط چند جمله‌ای $\bar{h}(x) = h_k + h_{k-1}x + \cdots + h_1 x^{k-1} + h_0 x^k$ است.

برهان. باید H از ابعاد $n(n-k)$ باشد. صرف نظر از درایه‌های $h_0, \cdots, h_k$، که تعداد آن‌ها $(k+1)$ تا است. در هر سطر $n-(k+1)$ تا صفر داریم. در هر مرحله h انتقال می‌دهیم پس $n-(k+1)$ انتقال صورت می گیرد که با سطر اول در کل $n-k$ داریم و سطرها نیز مستقل خطی هستند.

برای آن که ماتریس $n(n-k)$ ساخته شده، ماتریس مولد کد $C^\perp$ باشد؛ باید نشان دهیم سطرهای آن بر تمام کلمات کد عمود است؛ لذا کافی بر تمام سطرهای G عمود باشد چون کدواژه‌ها ترکیب سطرهای G هستند. فرض کنیم

$$G = \begin{bmatrix} g_0 & g_1 & \cdots g_r & 0 & 0 & \cdots 0 \\ 0 & g_0 & g_1 & \cdots g_r & 0 & \cdots 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots g_0 & g_1 & \cdots g_r \end{bmatrix}$$

باشد چون $h(x)$ چند جمله‌ای آزمون است داریم : در نتیجه $x^n - ۱ = h(x)g(x)$

$g(x) = g_۰ + g_۱x + \cdots + g_{n-k}x^{n-k} + \cdots + ۰x^{n-k+۱} + \cdots + ۰x^{n-۱}$. □

تعریف ۳.۱.۳. اگر R یک حلقه دلخواه باشد، در این صورت حلقه‌ی چندجمله‌ای روی R به صورت زیر تعریف می‌شود

$$R[x] = \left\{ \sum_{i=۰}^{m} a_i x^i | a_i \in R \right\}$$

اگر R حلقه دلخواه و θ اتومورفیسم از R باشد. در این صورت حلقه‌ی چندجمله‌ای اریب روی R به صورت:

$$R[x;\theta] = \left\{ \sum_{i=۰}^{m} a_i x^i | a_i \in R \right\}$$

$xa = \theta(a)x \quad \forall a \in R$. همچنین کدهای DNA با استفاده از کدهای دوری روی میدان $\mathbb{F}_۴$ تولید می‌شوند.

در مقالات [۴۷] و [۴۸] نویسندگان با استفاده از حلقه‌ی $\frac{\mathbb{F}_۲[u]}{(u^۲-۱)}$ کدهای DNA برگشت‌پذیر را بدست آوردند. در مقاله [۵۲] کدهای DNA دارای ۱۶ عنصر بررسی می‌شود.

مثال ۲.۱.۳. در حلقه‌ی $\frac{\mathbb{F}_۲[u]}{(u^۴-۱)}$ داریم:

$$u^۴ = ۱$$

$$u^۵ = u^۴u = u$$

$$u^۶ = u^۵u = u^۲$$

$$u^۷ = u^۶u = u^۳$$

$$\frac{\mathbb{F}[u]}{(u^۴ - ۱)} = \left\{ a_۰ + a_۱u + a_۲u^۲ + a_۳u^۳ | a_i \in \mathbb{F}_۲ \right\}$$

لذا دارای ۱۶ عنصر است.

در این پایان نامه با استفاده از کدهای دوری اریب، کدهای DNA برگشت‌پذیر را بررسی می‌کنیم

برای خاصیت برگشت‌پذیری DNA فرض کنید (a_1, a_2, a_3) یک کد واژه متناظر با رشته ACAGTC در DNA باشد، جایی که $a_1 \to AC$، $a_2 \to AG$، $a_3 \to TC$ در واقع هر دو نوکلئوتید را به یک حرف a نسبت دادیم . برگشت (a_1, a_2, a_3) به صورت (a_3, a_2, a_1) باشد و رشته TCAGAC ظاهرشود، اما در واقعیت، معکوس متناظر با TCAGAC رشته ACAGTC نمی‌باشد.

مولکول DNA که اغلب دستورالعمل ژنتیک را برای توابع سلولی حمل می‌کند، به صورت دو رشته‌ای می‌باشد و رشته‌ها DNA شامل ۴ نوکلئوتید است. نوکلئوتیدها دورشته‌ای را واتسون و کریک بیان کردند به طوری که A با T و C با G جفت می‌شوند. آدلمن[1] با الهام از مدل و اتسون و کریک که به اختصار (WCC) توانست؛ مسئله NP را حل کند با استفاده از چند جمله‌ای‌های θ ـ پالیندرومیک به خواص مقسوم علیه چندجمله‌ای $x^n - 1$ در حلقه چند جمله‌ای اریب $\mathbb{F}_{۴۲_s}[x;\theta]$ برای اولین بار می‌پردازیم. لازم است برخی خواص اساسی کد دوری اریب و کدهای DNA را بررسی کنیم.

تعریف ۴.۱.۳. فرض کنید C یک کد از طول n در $\mathbb{F}_q$ باشد. در این صورت برای هر $c = (c_0, c_1, \ldots, c_{n-1}) \in C$ داشته باشیم $c^r = (c_{n-1}, c_{n-2}, \ldots, c_1, c_0) \in C$ آن‌گاه C را کد برگشت‌پذیر می‌نامند. (کد برگشت‌پذیر که برعکس هر کدواژه نیز یک کدواژه از کد C باشد).

فرض کنید θ یک اتومورفیسم در میدان $\mathbb{F}_q$ باشد. در این صورت ،مجموعه چند جمله‌ای

$$\mathbb{F}_q[x;\theta] = \left\{a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} | a_i \in \mathbb{F}_q, n \in \mathbb{N}\right\}$$

جایی که عمل جمع در حلقه‌ی چند جمله‌ای اریب در میدان $\mathbb{F}_q$ مانند جمع چند جمله‌ای‌ها ، اما عمل ضرب به طور ویژه به صورت $x(a) = \theta(a)x$ لذا داریم $(a \in \mathbb{F}_q)$ (مرجع [۳۳] رابببنید)

تعریف ۵.۱.۳. کد دوری اریب، یک کد خطی C از طول n روی میدان $\mathbb{F}_q$ برای هر $(c_0, c_1, \ldots, c_{n-1}) \in C$ دارای خاصیت $(\theta(c_{n-1}), \theta(c_0), \ldots, \theta(c_{n-2})) \in C$ می‌باشد. (مرجع [۹] رابببنید.)

$$\varphi : \mathbb{F}q^n \to \frac{\mathbb{F}_q[x;\theta]}{(x_n - 1)}$$
$$c = (c_0, c_1, \ldots, c_{n-1}) \to c(x) = c_0 + c_1 x + \cdots + c_{n-1}x^{n-1}$$

[1] Adelman

در واقع برای نمایش چند جمله‌های یک کد از طول n، ما کدواژه‌های c را با عناصر $\varphi(c) = c(x)$ از $\frac{\mathbb{F}_q[x;\theta]}{(x^n-1)}$ متناظر می‌کنیم به طوری که ضرایب چندجمله‌ای مؤلفه کدواژه‌ها می‌شود. اگر m که مرتبه اتومورفیسم θ است، nرا عاد کند کد دوری اریب C از طول n در میدان $\mathbb{F}_q$ متناظر ایده‌آل چپ حلقه‌ی خارج قسمتی $\frac{\mathbb{F}_q[x;\theta]}{(x^n-1)}$ می‌باشد، اگر m، n عاد نکند $\frac{\mathbb{F}_q[x;\theta]}{(x^n-1)}$ حلقه نمی‌شود. در این صورت کد دوری اریب C به عنوان زیر مدول از $\frac{\mathbb{F}_q[x;\theta]}{(x^n-1)}$ می‌باشد. (مرجع [۴۶] رابببینید) در هر دو حالت کد C توسط یک چندجمله‌ای تکین $g(x)$ که مقسوم علیه راست از $x^n - 1$ در $\mathbb{F}_q[x;\theta]$ بدست می‌آید و داریم $C = \langle g(x) \rangle$. به علاوه اگر m و n نسبت به هم اول باشند، آن‌گاه C یک کد دوری روی میدان $\mathbb{F}_q$ است (مرجع [۴۶] را ببینید) به طور طبیعی کدهای دوری اریب کدهای خطی از طول n با بعد k و فاصله همینگ d روی میدان $\mathbb{F}_q$ هستند.

قضیه ۴.۱.۳. فرض کنید $\mathbb{F}$ یک میدان متناهی q عنصری باشد. در این صورت برای هر عنصر $\beta \in \mathbb{F}$ همواره داریم $\beta^q = \beta$.

اثبات:اگر $\beta = 0$ آن‌گاه حکم بدیهی است.

فرض کنید $\beta \neq 0$ باشد و عناصر ناصفر میدان $\mathbb{F}$ را β_1 و β_2 و ... و β_{q-1} نامگذاری کنیم یعنی

$$\mathbb{F}^* = (\beta_1, \beta_2, \ldots, \beta_{q-1})$$

عناصر ناصفر $\beta\beta_1$ و $\beta\beta_2$ و $\cdots$ و $\beta\beta_{q-1}$ را در نظر بگیرید. چون $\beta \neq 0$ و $\beta_i \neq 0$ برای هر $1 \leq i \leq q-1$، لذا در میدان داریم $\beta\beta_i \neq 0$ برای هر $1 \leq i \leq q-1$ در نتیجه در میدان $\mathbb{F}$ $\beta^{q-1} = 1$ و رابطه $\beta^q = \beta$ برقرار است.

تعریف ۶.۱.۳. عنصر α درمیدان $\mathbb{F}_q$ را مولد[1] نامند، هرگاه بتوان $\mathbb{F}_q$ را به فرم زیر نوشت $\mathbb{F}_q = \{0, \alpha, \alpha^2, \ldots, \alpha^{q-1}\}$.

تعریف ۷.۱.۳. مرتبه‌ی یک عنصر ناصفر $\alpha \in \mathbb{F}_q$، با نماد $\text{Order}(\alpha)$ نمایش داده می‌شود و برابر با کوچکترین عددصحیح مثبت K می‌باشد به گونه‌ای که $\alpha^K = 1$.

پژوهشگران چندجمله‌ای‌های ۴ شیفت شده و کدهای DNA برگشت‌پذیر را در میدان $\mathbb{F}_{16}$ بررسی می‌کنند. هر عناصر از $\mathbb{F}_{16}$ به صورت توان چهارم را به جفت‌های از DNA نگاشت دادند. جدول ۱.۳ تناظر نوکلئوتیدها را داریم به طوری که α^2 با جفت باز GC در DNA و α^8 با CG

[1] Generator

جدول ۱.۳: کدگذاری نوکلئوتیدها

AA	$\circ$	$\circ$
TT	α°	۱
AT	$\alpha^{۱}$	α
GC	$\alpha^{۲}$	$\alpha^{۲}$
AG	$\alpha^{۳}$	$\alpha^{۳}$
TA	$\alpha^{۴}$	$۱+\alpha$
CC	$\alpha^{۵}$	$\alpha+\alpha^{۲}$
AC	$\alpha^{۶}$	$\alpha^{۲}+\alpha^{۳}$
GT	$\alpha^{۷}$	$۱+\alpha+\alpha^{۳}$
CG	$\alpha^{۸}$	$۱+\alpha^{۲}$
CA	$\alpha^{۹}$	$\alpha+\alpha^{۳}$
GG	$\alpha^{۱\circ}$	$۱+\alpha+\alpha^{۲}$
CT	$\alpha^{۱۱}۱$	$\alpha+\alpha^{۲}+\alpha^{۳}$
GA	$\alpha^{۱۲}$	$۱+\alpha+\alpha^{۲}+\alpha^{۳}$
TG	$\alpha^{۱۳}$	$۱+\alpha^{۲}+\alpha^{۳}$
TC	$\alpha^{۱۴}$	$۱+\alpha^{۳}$

در [۳۸] یک الگوریتم از جدول ۴^s متناظر می‌کنیم با حالت تعمیم یافته $\mathbb{F}_{۴}^{۲s}$ و DNA با پایه $۲s$ با استفاده از چندجمله‌ای ۴^s شیفت آن‌ها می‌توان کدهای DNA برگشت‌پذیر را بدست آورد. بر طبق این الگوریتم نگاشت به‌صورت؛

$$\tau : \mathbb{F}_{۴^{۲s}} \to \{G, C, A, T\}^{۲s}$$

$$\beta \to (b_{\circ}, b_{۱}, \cdots b_{۲s-۱})$$

نگاشت طبیعی ϕ گسترش می‌یابد به نگاشت $\{G, C, A, T\}^{۲sn}$

$$\phi(c_{\circ}, c_{۱}, \ldots c_{c_{n-۱}}) = \big(\tau(c_{\circ})\tau(c_{۱}) \ldots \tau(c_{n-۱})\big)$$

$$\phi(c_{\circ}, c_{۱}, c_{۲}, c_{۳}) = (\alpha^{۲}, \alpha^{۲\circ}, \alpha^{۱۳۵}, \alpha^{۲۱۹})$$

جایی که برای هر $c_i \in \mathbb{F}_{۴^{۲s}}$ داریم $i \in \{\circ, \ldots, n-۱\}$. با توجه به جدول مرجع [۳۸] $\phi(c_{\circ}, c_{۱}, c_{۲}, c_{۳}) = (\alpha^{۲}, \alpha^{۲\circ}, \alpha^{۱۳۵}, \alpha^{۲۱۹}) = (AAAGATACCGACTAGA)$ در $\mathbb{F}_{۲۵۶}$ جایی α عنصر اولیه از $\mathbb{F}_{۲۵۶}$ می‌باشد. (تعریف ۶.۱.۳ را ببینید.)

تعریف ۸.۱.۳. فرض کنید $C \subseteq \mathbb{F}_{۴^{۲s}}^n$ باشد. اگر برای هر $c \in C$ داشته باشیم $\phi(c)^r \in \phi(C)$ آن‌گاه $\phi(C)$ کد DNA برگشت‌پذیر است.

در واقع تمام ۲^{sn} تایی‌هایی که روی $\{G, C, A, T\}$ اثر کرده باشد و برگشت آن هم در $\phi(C)$ باشد برای هر$c \in C$. طبق مرجع [۳۸] مشاهده می‌کنیم $\tau(\beta)$ و $\tau(\beta^{۴^{۲s}})$ معکوس هم هستند برای هر $\beta \in \mathbb{F}_{۴^{۲S}}$ (قضیه ۴.۱.۳ رابببینید.) در این صورت برگشت $\phi(c_۰, c_۱, \ldots, c_{n-۱})$ معادل $\phi(c_{n-۱}^{۴^s}, \ldots, c_۱^{۴^s}, c_۰^{۴^s})$ می‌توان گفت $\phi(c_۰, c_۱, \ldots, c_{n-۱}) = \phi(c_{n-۱}^{۴^s}, \ldots, c_۱^{۴^s}, c_۰^{۴^s})$، DNA برگشت‌پذیر برای همه می‌باشد. حلقه چندجمله‌ای اریب $\mathbb{F}_{۴^{۲s}}[x; \theta]$ برای هر $a \in \mathbb{F}_{۴^{۲s}}$ جایی که $\theta(a) = a^{۴^s}$.

در این پایان نامه در مورد ساختار برگشت‌پذیر DNA به وسیله کدهای دوری اریب در $\mathbb{F}_{۴^{۲s}}$ بحث می‌کنیم. یکسری خواص اضافی تر را روی مولد کد دوری اریب در میدان $\mathbb{F}_۴^{۲s}$ تحمیل می‌کنیم. همچنین نشان می‌دهیم چندجمله‌ای خودمتقابل اریب نیز یک θ ـ پالیندرومیک است، دوگان کدهای دوری اریب نیز کدهای DNA برگشت‌پذیر است.

۲.۳ دوگان θـ کدهای دوری

لم ۱.۲.۳. فرض کنید که مرتبه‌ی θ ، n را عاد کند. همچنین فرض کنید: $X^n - ۱ = h.g \in Z(\mathbb{F}_q[X, \theta])$. و کد C یک θـ کد دوری متناظر با ایده‌آل چپ تولید شده توسط g در حلقه خارج‌قسمتی $\frac{\mathbb{F}_q[X;\theta]}{(X^n-۱)}$ باشد. اگر داشته باشیم:

$$g = g_۰ + g_۱X + \cdots + g_rX^r, \quad h = h_۰ + h_۱X + \cdots + h_{n-r}X^{n-r},$$

آنگاه ماتریس زیر ماتریس کنترل توازن برای کد C، می‌باشد.

$$H = \begin{bmatrix} h_{n-r} & \cdots & \theta^{n-r-۱}(h_۰) & \theta^{n-r}(h_۰) & ۰ & \cdots & ۰ \\ ۰ & \theta(h_{n-r}) & \cdots & \cdots & \theta^{n-r+۱}(h_۰) & \cdots & ۰ \\ ۰ & \ddots & \ddots & \cdots & \cdots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \cdots & \cdots & \ddots & ۰ \\ ۰ & \cdots & ۰ & \theta^{r-۱}(h_{n-r}) & \cdots & \theta^{n-۲}(h_۱) & \theta^{n-۱}(h_۰) \end{bmatrix}$$

برهان. اگر $a(X) \in C$ باشد، آن‌گاه در حلقه خارج قسمتی $\frac{\mathbb{F}_q[X;\theta]}{(X^n-1)}$، و $a(X).h = 0$. با توجه به اینکه درجه $(a(X).h)$، کوچکتر از $n+k$ می‌باشد، که در آن $k = n-r$ بعد کد C و درجه‌ی h است. می‌توان نتیجه گرفت که ضرایب $X^k, X^{k+1}, \ldots, X^{n-1}$ باید صفر باشد. همچنین برای $l \in \{0, \ldots, r-1\}$، ضرایب $a(X).h\ X^{l+k}$ برابر با $\sum_{i=1}^{l+k} a_i\theta^i(h_{l+k}-i)$ می‌باشد. لذا $a(X) \in C$ است، نتیجه می‌گیریم $H.a^t = 0$. بنابراین کد تولید شده توسط ماتریس H، مشمول در دوگان کد C $(C^\perp)$ می‌باشد. از آنجا که بعد کد $C^\perp$ برابر با $n - dim(C) = r$، تعداد سطرهای H است نتیجه می‌گیریم که H یک ماتریس مولد برای دوگان کد C می‌باشد. □

نتیجه ۱.۲.۳. فرض کنید که مرتبه خودریختی θ، n را عاد کند. و همچنین فرض کنید که $g = \sum_{i=0}^{n-r} g_iX^i$ و $h = \sum_{i=0}^{n-r} h_iX^i$ ، عناصری از حلقه $\mathbb{F}_q[x;\theta]$ باشند، به طوری که

$$h.g = X^n - 1 \in Z(\mathbb{F}_q[X,\theta])$$

باشد. دوگان θ ـ کد دوری تولید شده توسط g در حلقه‌ی خارج قسمتی $\frac{\mathbb{F}_q[X;\theta]}{(X^n-1)}$، θ ـکد دوری تولید شده توسط $g^\perp = h_{n-r} + \theta(h_{n-r-1})X + \cdots + \theta^{n-r}X(h_0)X^{n-r}$ می‌باشد.

برهان. همانطور که می‌دانیم ماتریس کنترل توازن H در یک کد، یک ماتریس مولد برای کد دوگان می‌باشد. اکنون باید نشان دهیم که H، ماتریسی از یک θ ـ کد دوری می‌باشد. بر اساس نتیجه قبل، $\theta^{n-r}(h_0)X^{n-r} + \cdots + \theta(h_{n-r-1})X + h_{n-r}$ یک مقسوم علیه از $X^n - 1$ می‌باشد.

اما از آنجا که حلقه‌ی $\mathbb{F}_q[X,\theta]$ یک دامنه‌ی اور راست است، دارای یک میدان کسر اریب راست است. فرض کنید که $\mathbb{F}_q[X,\theta]$ نمایش دهنده‌ی میدان کسر راست از حلقه $\mathbb{F}_q[X,\theta]$ باشد و همچنین فرض کنید X^{-1} معکوس X باشد. در این صورت داریم $aX^{-1} = X^{-1}\theta(a)$. حال فرض کنید که حلقه‌ی $R \subset \mathbb{F}_q(X,\theta)$، شامل عناصر $\sum_{i=0}^{n} X^{-i}a_i$ باشد که در آن ضرایب در سمت راست قرار گرفته و قانون ضرب به صورت $aX^{-1} = X^{-1}\theta(a)$ می باشد. لذا حلقه‌ی R یکریخت است با حلقه‌ی چندجمله‌ای‌های اریب $\mathbb{F}_q[X^{-1},\theta^{-1}]$. نگاشت:

$$\phi : \mathbb{F}_q(X,\theta) \to R \subset \mathbb{F}_q(X,\theta)$$

$$\sum_{i=0}^{n} a_iX^i \to \sum_{i=0}^{n} X^{-i}a_i$$

یک یکریختی حلقه‌ای نمی‌باشد. لذا برای $P_۱ = \sum_{i=۰}^{s} a_i X^i$ و $P_۲ = \sum_{i=۰}^{t} b_i X^i$ داریم:

$$\begin{aligned}
\varphi(P_۱ + P_۲) &= \varphi(P_۱ + P_۲) \\
\varphi(P_۱P_۲) &= \varphi(\sum_{k=۰}^{s+t} \sum_{i+j=k} a_i X^i b_j X^j) \\
&= \varphi(\sum_{k=۰}^{s+t} \sum_{i+j=k} a_i.\theta^i(b_j) X^{i+j} \\
&= (\sum_{k=۰}^{s+t} X^{-k} \sum_{i+j=k} \theta^i(b_j).a_i \\
&= (\sum_{k=۰}^{s+t} \sum_{i+j=k} X^{(}-۱)(X^{-}i\theta^i(b_j))a_i \\
&= (\sum_{k=۰}^{s+t} \sum_{i+j=k} X^{-j} b_j X^{-i} a_i) \\
&= \varphi(P_۱) = \varphi(P_۲).
\end{aligned}$$

اگر $X^n - ۱$ باشد آن‌گاه داریم:

$$\varphi(X^n - ۱) = X^{-n} - ۱ = \varphi(h.g) = \varphi(g.h) = (\sum_{j=۰}^{n-r} X^{-j} h_j)(\sum_{k=i=۰}^{r} X^{-i} g_j).$$

به این ترتیب در میدان $\mathbb{F}_q[X, \theta]$ داریم:

$$X^{n-r}(X^{-n} - ۱)X^r = -(X^n - ۱)$$

$$= (h_{n-r} + \theta(h_{n-r-۱})X + \cdots + \theta^{n-r}(h_۰)X^{n-r}).\bar{g}$$

که در آن $\bar{g} \in \mathbb{F}_q[X, \theta]$ می‌باشد. بنابراین چندجمله‌ای $(h_{n-r} + \theta(h_{n-r-۱})X + \cdots + \theta^{n-r}(h_۰)$ چندجمله‌ای $X^n - ۱$ ، در حلقه‌ی $\mathbb{F}_q[X, \theta]$ را عاد می‌کند. □

مثال ۱.۲.۳. حلقه‌ی چندجمله‌ای‌های اریب $\mathbb{F}_۴[X, \theta]$ را در نظر بگیرید، که در آن θ خودریختی فروبنیوس و α به عنوان مولد $\mathbb{F}_۴^*$ باشد . با استفاده از چندجمله‌ای‌های

$$g = X^۲ + \alpha X + \alpha^۲$$

و

$$h = X^۲ + \alpha X + \alpha^۲$$

در حلقه‌ی $\mathbb{F}_۴[X, \theta]$ نتیجه می‌گیریم:

$$h.g = X^۴ - ۱$$

و چندجمله‌ای

$$g^{\perp} = \theta^۲(\alpha)X^۲ + \theta(\alpha)X + ۱ = \alpha X^۲ + \alpha^۲ X + ۱$$

خواهد بود. همچنین با استفاده از $\alpha^۲ g^{\perp} = g$ نتیجه می‌گیریم که کد تولید شده توسط g در حلقه‌ی خارج قسمتی $\frac{\mathbb{F}_۴[x;\theta]}{(X^۴-۱)}$ ، یک کد خوددوگان اقلیدسی می‌باشد.

مثال ۲.۲.۳. فرض α یک عنصر اولیه از میدان $\mathbb{F}_{۱۶}$ باشد داریم $f(x) = ۱ + \alpha x + \alpha^۲ x^۲ + \alpha^۴ x^۳ + x^۴ \in \mathbb{F}_{۱۶}[x; \theta]$

جایی $\theta(a) = a^۴$ بنابراین $f(x) = \mathbb{F}^R(x)$ اما از آنجایی $\theta(\alpha^۲) \neq \alpha^۲$ و $f(x)$ یک چند جمله‌ای غیر پالیندرومیک است.

$$f(x) = ۱ + \alpha x + \alpha^۲ x^۲ + \alpha^۴ x^۳ + x^۴ \in \mathbb{F}_{۱۶}[x; \theta]$$
$$f^R(x) = \sum_{i=۰}^{t} x^i a_{t-i} = \sum_{i=۰}^{t} \theta^i(a_{t-i}) x^i$$

قضیه ۱.۲.۳. فرض کنید $\mathbb{F}$ یک میدان متناهی از مشخصه P باشد و θ یک اتومورفیسم $\mathbb{F}$ باشد. یک زیر مجموعه Cاز $\mathbb{F}^۲$ رایک کددوری اریب از طول n می نامند هرگاه

الف) C یک زیر فضایی از $\mathbb{F}^n$ باشد.

ب) اگر $c = (c_۰, c - ۱, \cdots, c_{n-۱}) \in C$ آن‌گاه $\theta(c) = \theta(c_{n-۱}), \theta(c - ۰), \cdots, \theta(c_{n-۲}) \in C$ لذا در کد دوری اریب جمع مانند جمع معمولی می‌باشد ولی در مورد ضرب آنها اگر $m|n$ این صورت ضرب X در کدهای دوری اریب ایده‌آل $\frac{\mathbb{F}_q[x;\theta]}{(x^n-۱)}$ هستند.

تعریف ۱.۲.۳. فرض کنید $f(x) = a_۰ + a_۱ x + \cdots + a_t x^t$ یک چند جمله‌ای با درجه‌ی t بر روی میدان $\mathbb{F}_q$ باشد و θ یک اتومورفیسم از $\mathbb{F}_q$ برای هر $i \in \{۰, ۱, \ldots, t\}$ اگر $a_i = a_{t-i}$ باشد آن‌گاه

$f(x)$ یک چند جمله‌ای پالیندرومیک می‌باشد و برای هر $i \in \{0, 1, \ldots, t\}$ اگر $a_i = \theta a_{(t-i)}$ باشد آن‌گاه $f(x)$ یک چند جمله‌ای θ ـ پالیندرومیک می‌باشد.

تعریف ۲.۲.۳. چند جمله‌ای متقابل $f(x) = \sum_{i=0}^{t} a_i x^i \in \mathbb{F}[x, \theta]$ با درجه t به صورت $f^R(x) = \sum_{i=0}^{t} x^i a_{t-i} = \sum_{i=0}^{t} \theta^i (a_{t-h}) x^i$ تعریف می شود. در این صورت اگر $f(x) = \mathbb{F}^R(x)$ آن‌گاه $f(x)$ چند جمله‌ای متقابل اریب تعریف می‌شود.

برهان. فرض کنید درجه θ، n عاد کند. در این صورت اگر $x^n - 1 = h(x)g(x) \in \mathbb{F}[x; \theta]$ باشد. کد C کد دوری اریب از طول n در $\mathbb{F}_q$ با چند جمله‌ای $g(x)$ تولید شود. در این صورت دوگان کد C یک کد دوری اریب از طول n با $h^R(x)$ تولید می‌شود یعنی $C^{\perp} = \langle h^R(x) \rangle$. □

۳.۳ کدهای DNA برگشت‌پذیر

فرض حلقه‌ی چند جمله‌ای اریب $\mathbb{F}_q[x, \theta]$ با اتومورفیسم θ در $\mathbb{F}_q$ باشد، جایی که $q = 4^{2s}$ لذا داریم $\theta(a) = a^{4^s}$ چون مرتبه θ برابر ۲ است؛ کد دوری اریب از طول فرد در $\mathbb{F}_q$ نداریم و کد دوری معمولی می‌باشد. [۴۶] لازم است برای درک بهتر؛ کدهای DNA برگشت‌پذیر از طول زوج و همچنین طول فرد را جداگانه بررسی می‌کنیم.

۴.۳ کدهای DNA با طول زوج

در این بخش مقسوم علیه‌های $x^n - 1$ جایی که n زوج باشد را داریم. فرض کنید C یک کد دوری اریب از طول n روی میدان $\mathbb{F}_q$ با اتومورفسیم θ باشد. $g(x)$ یک چند جمله‌ای تکین، ناصفر از کمترین درجه در C باشد آن‌گاه، C توسط $g(x)$ تولید می‌شود علاوه بر این $g(x)$ مقسوم علیه راست از $x^n - 1$ در $\mathbb{F}_q[x, \theta]$ (لم[۹]رابببینید) می‌توان گفت $\langle g(x) \rangle$ ایده‌آل تولید شده $\langle \beta g(x) \rangle$ جایی که $\beta \in \mathbb{F}_q^*$ برابر $\mathbb{F}_q[x; \theta]$ می‌باشد. اگر $x^n - 1 = h(x)g(x)$، آن‌گاه $x^n - 1 = \left(h(x)\beta^{-1}\right)(\beta g(x))$ در $\mathbb{F}_q[x; \theta]$. در این صورت اگر $g(x)$ مقسوم علیه راست از $x^n - 1$ داریم $\beta g(x)$. پس چند جمله‌ای $g(x) = g_0 + g_1 x + \cdots + g_m x^m$ را به عنوان یک چندجمله‌ای از C با کمترین درجه در C بدون این که تکین باشد.

قضیه ۱.۴.۳. فرض کنید $g(x)$ یک مقسوم‌علیه راست از $x^n - ۱$ در $\mathbb{F}_q[x;\theta]$ باشد، جایی که درجه $g(x) = m$ عددی زوج باشد. آن‌گاه کد دوری اریب تولید شده توسط $C = \langle g(x) \rangle$ در $\mathbb{F}_q$ با طول n یک کد DNA برگشت‌پذیر می‌باشد اگر و تنها اگر $g(x)$ یک چندجمله‌ای پالیندرومیک باشد.

برهان. فرض کنید $g(x)$ یک چندجمله‌ای پالیندرومیک و ϕ متناظر کدواژه‌ها در DNA باشد. معکوس هر کدواژه DNA $\phi(c)$ برای هر $c \in C$ توسط معادلهٔ زیر به‌دست می‌آید:

$$\left(\phi\left(\sum_{i=۰}^{k-۱}\beta_i x^i g(x)\right)\right)^r = \phi\left(\sum_{i=۰}^{k-۱}\theta(\beta_i)x^{k-۱-i}g(x)\right)$$

جایی‌که $k = n - deg(g(x))$ و $\beta_i \in \mathbb{F}_q$ چون $\sum_i \theta(\beta_i)x^{k-۱-i}g(x) \in C$ آن‌گاه C یک کد DNA برگشت‌پذیر می‌باشد.

$$\sum_{i=۰}^{k-۱} a_i x^i$$

$$a_۰ + a_۱x + \cdots + a_{k-۲}x^{k-۲} + a_{k-۱}x^{k-۱}$$

$$\sum_{i=۰}^{k-۱} a_i x^{(k-۱)-i}$$

$$a_۰x^{k-۱} + a_۱x^{k-۲} + \cdots + a_{k-۲}x + a_{k-۱}$$

ضرایب به صورت زیر می‌باشد.

$$(a_۰, a_۱, a_۲, \ldots, a_{k-۱})$$

$$(a_{k-۱}, a_{k-۲}, \ldots, a_۲, a_۱, a_۰)$$

. برعکس، فرض کنید $C = \langle g(x) \rangle$ یک کد DNA برگشت‌پذیر باشد. ثابت کنیم $g(x)$ پالیندرومیک است. فرض کنید:

$$a(x) = g_m^{-۱}g(x) = a_۰ + a_۱x + \cdots + a_{m-۱}x^{m-۱} + x^m$$

نکته :چون ضرایب از میدان $\mathbb{F}_q[x;\theta]$ می‌باشد و $deg(g(x)) = m$ است. $g_m \neq ۰$ بنابراین $g_m \in \mathbb{F}_q$ لذا $g_m^{-۱}$ وجود دارد. جایی که $a_i = g_m^{-۱}g_i$. آن‌گاه $a(x) \in C$ یک چندجمله‌ای ناصفر و

مونیک از کمترین درجه در C می‌باشد (با توجه به انتخاب $a(x)$ که از $g(x)$ به‌دست آمده است و قطعاً در کد می‌افتد و به ناصفر بودن x^m توجه کنید)
جایی‌که

$$c_1(x) = x^{n-m-1}a(x)$$
$$= x^{n-m-1}(a_0 + a_1x + \cdots + a_{m-1}x^{m-1} + x^m)$$
$$= \theta^{n-m-1}(a_0)x^{n-m-1} + \theta^{n-m-1}(a_1)x^{n-m-1} + \cdots + \theta^{n-m-1}(a_{m-1})x^{n-2} + x^{n-1}$$

$ord(\theta) = 2$ بنابراین داریم:

$$= \theta(a_0)x^{n-m-1} + \theta(a_1)x^{n-m} + \cdots + \theta(a_{m-1})x^{n-2} + x^{n-1}$$
$$= a_0^{4^s}x^{n-m-1} + a_1^{4^s}x^{n-m} + \cdots + a_{m-1}^{4^s}x^{n-2} + x^{n-1} \in C$$

چون $a(x) \in C$ و $c_1(x) = x^{n-m-1}a(x)$ لذا $c_1(x) \in C$. با توجه به کد DNA برگشت‌پذیر $c_1(x) \in C$ در این صورت $c_2(x)$ که برگشت $c_1(x)$ در کد C قرار می‌گیرد و کد دوری اریب می‌باشد.

در این پایان‌نامه نوع خاص از اُتومورفیسم $\theta : \mathbb{F}_q \to \mathbb{F}_q$ داریم به‌طوری که اتومورفیسم جنس DNA می‌پذیرد.

$$\forall a \in \mathbb{F}_q \;\; \theta(a) = a^{4^s}$$

بررسی می‌کنیم که $\theta^{n-m-1} = \theta(a_0)$ و با توجه به میدان qـ عنصری و فرض مرتبهٔ اُتومورفیسم برابر ۲ می‌باشد. $c_2(x)$ عکس $c_1(x)$ (دارای 4^s است) کد برگشت‌پذیر DNA می‌باشد به‌صورت :

$$c_2(x) = 1 + a_{m-1}x + \cdots + a_1x^{m-1} + a_0x^m \in C$$

. اکنون $c_3(x)$ را می‌سازیم:

$$c_3(x) = a(x) - a_0^{-1}c_2(x)$$
$$= (a_0 + a_1x + \cdots + a_{m-1}x^{m-1} + x^m) - (a_0^{-1} + a_0^{-1}a_{m-1}x + \cdots + a_0^{-1}a_1x^{m-1} + x^m)$$
$$= (a_0 - a_0^{-1}) + (a_0 - a_0^{-1}a_{m-1}) + \cdots + (a_{m-1} - a_0^{-1}a_1)x^{m-1}$$

چون $deg(c_۳(x)) < deg(g(x))$ و $c_۳(x) \in C$، لذا باید $c_۳(x) = ۰$ باشد لذا داریم:

$$a_۰ - a_۰^{-۱} = ۰ \quad \rightarrow \quad a_۰ = a_۰^{-۱} \rightarrow \quad a_۰ = ۱$$

$$a_۱ - a_۰^{-۱} a_{m-۱} = ۰ \quad \rightarrow \quad a_۱ - a_{m-۱} = ۰ \rightarrow \quad a_۱ = a_{m-۱}$$

$$\vdots$$

$$a_{m-۱} - a_۰^{-۱} a_۱ = ۰ \quad \rightarrow \quad a_{m-۱} - a_۱ = ۰ \rightarrow \quad a_{m-۱} = a_۱$$

$a_{m-۱} = a_۱$ ، چون $a(x)$ پالیندرومیک می‌باشد آن‌گاه $g(x) = g_m a(x)$ پالیندرومیک می‌باشد.

□

قضیه ۲.۴.۳. فرض کنید $g(x)$ یک مقسوم‌علیه راست از $x^n - ۱$ در $\mathbb{F}_q[x;\theta]$، جایی که درجه $g(x) = m$ فرد باشد. $C = \langle g(x) \rangle$ یک کد دوری اریب از طول n روی میدان $\mathbb{F}_q$ باشد. در این صورت اگر $g(x)$ یک چندجمله‌ای θـ پالیندرومیک باشد. آنگاه، C یک کد DNA برگشت‌پذیر می‌باشد و بالعکس C یک کد DNA برگشت‌پذیر باشد آن‌گاه C توسط یک چندجمله‌ای θـ پالیندرومیک تولید می‌شود

برهان. فرض کنید $g(x)$ یک چندجمله‌ای θـ پالیندرومیک باشد. ϕ متناظر کدواژه‌ها در DNA باشد. معکوس هر کدواژه DNA، $\phi(c)$ برای هر $c \in C$ ،توسط معادلهٔ زیر به‌دست می‌آید:

$$\left(\phi \left(\sum_{i=۰}^{k-۱} \beta_i x^i g(x) \right) \right)^r = \phi \left(\sum_{i=۰}^{k-۱} \theta(\beta_i) x^{k-۱-i} g(x) \right)$$

جایی که $k = n - deg(g(x))$ و $\beta_i \in \mathbb{F}_q$ چون $\sum_{i=۰}^{k-۱} \theta(\beta_i) x^{k-۱-i} g(x) \in C$ آن‌گاه C یک کد DNA برگشت‌پذیر می‌باشد.

برعکس، فرض کنید $C = \langle g(x) \rangle$ یک کد DNA برگشت‌پذیر باشد.

$$a(x) = g_m^{-۱} g(x) = a_۰ + a_۱ x + \cdots + a_{m-۱} x^{m-۱} + x^m$$

جایی که $a_i = g_m^{-۱} g_i$ بنابراین $a(x) \in C$ یک چندجمله‌ای ناصفر و مونیک از کمترین درجه در C می‌باشد (با توجه به انتخاب $a(x)$ که از $g(x)$ به‌دست آمده است و قطعاً در کد می‌افتد و به ناصفر بودن x^m توجه کنید)

$$c_۱(x) = x^{n-m-۱} a(x) = x^{n-m-۱} (a_۰ + a_۱ x + \cdots + a_{m-۱} x^{m-۱} + x^m)$$

تذکر: چون $n-m-۱$ عددی زوج است با توجه به زوج بودن n و فرد بودن m کد DNA برگشت‌پذیر به صورت

$$c_۲(x)=۱+a_{m-۱}^{۴^s}x+\cdots+a_۱^{۴^s}x^{m-۱}+a_۰^{۴^s}x^m\in C.$$

(نکته ۴.۳ مراجعه شود) چون $c_۲(x)\in C$، کد DNA برگشت‌پذیر می‌باشد. $c_۳(x)$ را می سازیم:

$$c_۳(x)=a(x)-a_۰^{-۴^s}c_۲(x)=(a_۰-a_۰^{-۴^s})+(a_۱-a_۰^{-۴^s}a_{m-۱}^{-۴^s})x+\cdots+(a_{m-۱}-a_۰^{-۴^s}a_۱)x^{m-۱}\in C.$$

درجه $c_۳(x)$ کمتر از درجه $g(x)=m$ می شود زیرا ضریب پیشرو از بین رفته است و با توجه به $g(x)$ مولد می‌باشد و درجه $c_۳(x)$ حداکثر $m-۱$ می‌باشد، امکان ندارد مگر این که $c_۳(x)=۰$ باشد.

$$c_۳(x)=۰\to a_۰-a_۰^{-۴^s}=۰\to a_۰^{۴^s+۱}=۱\to a_۰=\alpha^{(۴^s-۱)j}.$$

$a_۰$ برحسب عنصر اولیه α می‌نویسم. اگر α در $\mathbb{F}_q$ باشد و j یک عدد صحیح $c_۳(x)=a(x)-a_۰^{-۴^s}c_۲(x)=(a_۰-a_۰^{-۴^s})+(a_۱-a_۰^{-۴^s}a_{m-۱}^{-۴^s})x+$ باشد

$$a(x)=\sum_{i=۰}^{\frac{m-۱}{۲}}\left(a_ix^i+\alpha^{(۴^s-۱)j}a_i^{۴^s}x^{m-i}\right)$$

جایی که، $a_۰=\alpha^{(۴^s-۱)j}$ اگر از سمت چپ $a(x)$ را در α^j ضرب کنیم به صورت:

$$\alpha^ja(x)=\sum_{i=۰}^{\frac{m-۱}{۲}}\alpha^ja_ix^i+\alpha^{(۴^s)j}a_i^{۴^s}x^{(m-i)}\in C$$

بنابراین $\alpha^ja(x)=\alpha^jg_m^{-۱}g(x)$ یک چندجمله‌ای θ ـ پالیندرومیک در C است.

$$\langle\alpha^jg_m^{-۱}g(x)\rangle=\langle g(x)\rangle=C$$

می توان نتیجه گرفت C یک چند جمله‌ای θ ـ پالیندرومیک است. □

لم ۱.۴.۳. اگر $h.g\in\mathbb{F}_q[X,\theta])$باشد بنابراین $h.g=g.h\in\mathbb{F}_q[X,\theta]$.

برهان. فرض $h.g\in\mathbb{F}_q[X,\theta])$ باشد؛ ما $h.(g.h)=(h.g).h=h.(h.g)$. پس با توجه به لم ۳.۳.۱ مرجع [۱۱] لم زیر را می‌توان نتیجه گرفت. □

لم ۲.۴.۳. فرض کنید $x^n-۱=h(x)g(x)$ در $\mathbb{F}_q[x;\theta]$. در این صورت $x^n-۱=g(x)h(x)$ یعنی مقسوم علیه راست $x^n-۱$ یک مقسوم علیه چپ در $\mathbb{F}_q[x;\theta]$ می‌باشد.

قضیه ۳.۴.۳. فرض کنید $x^n-۱=h(x)g(x)$ در $\mathbb{F}_q[x;\theta]$ جایی که درجه‌ی چند جمله‌ای $g(x)$ زوج می‌باشد. در این صورت اگر $h(x)$ چند جمله‌ای پالیندرومیک باشد، آن‌گاه $g(x)$ نیز چند جمله پالیندرومیک می‌باشد.

برهان. فرض کنید $h(x)=h_۰+h_۱x+\cdots+h_{۲k}x^{۲k}$ و $g(x)=g_۰+g_۱x+\cdots+g_{۲r}x^{۲r}$، جایی که $n=۲r+۲k$. اگر $h(x)$ چند جمله‌ای پالیندرومیک باشد، آن‌گاه برای هر $i=۰,۱,\ldots,k$ برابر $h_i=h_{۲k-i}$. فرض کنید a_i ضریب x^i در $h(x)g(x)$ باشد برای هر $t<\frac{n}{۲}$ ضریب x^t در $h(x)g(x)$ برابر است با $a_t=\sum_{j=۰}^{t}h_j\theta^j(g_{t-j})$ و ضریب x^{n-t} به صورت $a_{n-t}=h_i=h_{۲k-i}$ $\sum_{j=۰}^{t}h_{۲k-j}\theta^{۲k-j}\left(g_{۲r-(t-j)}\right)$ است.

تساوی $h(x)g(x)=x^n-۱$ نتیجه می‌دهد اگر $a_۰=a_n=۱$ و $a_i=۰$ با استفاده از استقرار نشان می‌دهیم که برای هر $i=۱,\ldots,n-۱$ داریم $g_i=g_{۲r-i};i=۰,۱,\ldots,r$.

$a_۰=a_n=۱$ چون $a_n=h_{۲k}\theta^{۲k}(g_{۲r})=h_{۲k}g_{۲r}$ از سوی دیگر $a_۰=h_۰\theta^{۰}(g_۰)=h_۰g_۰\ ;i=۰$ و $h_۰=h_{۲k}$ بنابراین داریم: $g_۰=g_{۲r}$.

فرض کنید (فرض استقرا) که تساوی $g_i=g_{۲r-i}$ برای هر i که $۰<i<l$ برقرار باشد. (جایی که $l<r$) .

در این صورت ضرایب a_l و a_{n-l} را در نظر بگیریم:

$$a_l=\sum_{j=۰}^{l}h_j\theta^j(g_{l-j})=\sum_{j=۱}^{l}h_j\theta^j(g_{l-j})+h_۰g_l$$
$$a_{n-l}=\sum_{j=۰}^{l}h_{۲k-j}\theta^{۲k-j}(g_{۲r-(l-j)})=\sum_{j=۱}^{۱}h_{۲k-j}\theta^{۲k-j}(g_{۲r-(l-j)})+h_{۲k}g_{۲r-l}$$

چون مرتبه θ برابر ۲ می‌باشد، در این صورت برای هر $a\in\mathbb{F}_q$ و هر $j\in\{۱,\ldots l\}$ داریم $\theta^j(a)=\theta^{۲k-j}(a)$ چون $h_j=h_{۲k-j}$ و $g_{l-j}=g_{۲r-(l-j)}$، لذا داریم:

$$h_j\theta^j(g_{l-j})=h_{۲k-j}\theta^{۲k-j}\left(g_{۲r-(l-j)}\right)$$

بنابراین $a_l=a_{n-l}=۰$ چون $\sum_{j=۱}^{l}h_j\theta^j(g_l-j)=\sum_{j=۱}^{l}h_{۲k-j}\theta^{۲k-j}\left(g_{۲r-(l-j)}\right)$ می‌توان

نتیجه گرفت که $h_\circ g_l = h_{۲k} g_{۲r-l} = h_\circ g_{۲r-l}$؛ لذا $g_l = g_{۲r-۱}$. بنابراین $g(x)$ یک چندجمله‌ای پالیندرومیک می‌شود. □

با استفاده از لم ۲.۴.۳ و قضیه۳.۴.۳ نتیجه‌ی زیر حاصل می‌شود.

نتیجه ۱.۴.۳. فرض کنید $x^n - ۱ = h(x)g(x)$ در $\mathbb{F}_q[x;\theta]$، جایی که درجه‌ی چندجمله‌ای $g(x)$ زوج می‌باشد. اگر $g(x)$ یک چند جمله‌ای پالیندرومیک باشد، آن‌گاه $h(x)$ نیز یک چند جمله‌ای پالیندرومیک می‌باشد.

قضیه ۴.۴.۳. فرض کنید $x^n - ۱ = h(x)g(x)$ در $\mathbb{F}_q[x;\theta]$، جایی که درجه‌ی چند جمله‌ای $g(x)$ فرد می‌باشد. در این صورت اگر $g(x)$ یک چند جمله‌ای θ ـ پالیندرومیک باشد آن‌گاه $h(x)$ یک چند جمله‌ای پالیندرومیک می‌باشد.

برهان. فرض کنید $h(x) = h_\circ + h_۱x + \cdots + h_{۲k-۱}x^{۲k-۱}$ و $g(x) = g_\circ + g_۱x + \cdots + g_{۲r-۱}x^{۲r-۱}$ جایی که $n = ۲r + ۲k - ۲$. فرض کنید $g(x)$ چند جمله ای θ ـ پالیندرومیک باشد برای هر $i = \circ, ۱, \ldots, r-۱$. برابر $g_i = \theta(g_{۲r-۱-i})$. آن‌گاه a_i ضریب x^i در $h(x)g(x)$ باشد. برای هر $t < \frac{n}{۲}$ ضریب x^t $h(x)g(x)$ به صورت $a_t = \sum_{j=\circ}^{t} h_j\theta^j(g_{t-j})$. و ضریب x^{n-t} برابر است با $a_{n-t} = \sum_{j=\circ}^{t} h_{۲k-۱-j}\theta^{۲k-۱-j}(g_{۲r-۱-(t-j)})$.

تساوی $x^n - ۱ = g(x)h(x)$ نتیجه می‌دهد $a_\circ = a_n = ۱$ و $a_i = \circ$ برای هر $i = ۱, \cdots, n-۱$.

با استفاده از استقرا نشان می‌دهیم که برای هر $i = \circ, ۱, \cdots, k-۱$ داریم $h_i = h_{۲k-۱-i}$ برای $i = \circ$ داریم $a_\circ = h_\circ\theta^\circ(g_\circ) = h_\circ g_\circ$ ازسوی دیگر $a_n = h_{۲k-۱}\theta^{۲k-۱}(g_{۲r-۱}) = h_{۲k-۱}\theta(g_{۲r-۱})$..

چون $a_\circ = a_n = ۱$ و $g_\circ = \theta(g_{۲r-۱})$ لذا داریم $h_\circ = h_{۲k-۱}$.

فرض کنید استقرا که تساوی $h_i = h_{۲k-۱-i}$ برای هر $\circ < i < l$ برقرار باشد. جایی که $l \le k-۱$. در این صورت ضرایب a_l تا a_{n-l} را در نظر بگیریم:

$$a_l = \sum_{j=\circ}^{l} h_j\theta^j(g_{l-j}) = \sum_{j=\circ}^{l-۱} h_j\theta^j(g_{l-j}) + h_l\theta^l(g_\circ)$$

$$a_{n-l} = \sum_{j=\circ}^{l} h_{۲k-۱-j}\theta^{۲k-۱-j}g_(۲r-۱-(l-j))$$

$$= \sum_{j=\circ}^{l-۱} h_{۲k-۱-j}\theta^{۲k-۱-j)}g_{۲r-۱-(l-j)} + h_{۲k-۱-l}\theta^{۲k-۱-l}(g_{۲r-۱}).$$

چون مرتبه θ برابر ۲ می‌باشد، در این صورت برای هر $a \in \mathbb{F}_q$ داریم $\theta^j(a) = \theta^{2k-1-j}(\theta(a))$ و $j \in \{0, \ldots, l-1\}$ چون $h_j = h_{2k-1-j}$ و $g_{l-j} = \theta(g_{2r-(l-j)})$ داریم:

$$\theta^j(g_{l-j}) = \theta^{2k-1-j}(\theta(g_{l-j})) = \theta^{2k-1-j}(g_{2r-1-(l-j)})$$

همچنین $h_j\theta^j(g_{l-j}) = h_{2k-1-j}\theta^{2k-1-j}(g_{2r-1-(l-j)})$. لذا داریم:

$$\sum_{j=0}^{l-1} h_j\theta^j(g_{l-j}) = \sum_{j=0}^{l-1} h_{2k-1-j}\theta^{2k-1-j}(g_{2r-1-(l-j)})$$

چون $a_l = a_{n-l} = 0$، می‌توان نتیجه گرفت که $h_1 g_0 = h_{2k-1-l}\theta(g_{2r-1}) = h_1$. بنابراین $h(x)$ یک چند جمله‌ای پالیندرومیک می‌باشد. □

قضیه ۵.۴.۳. فرض کنید $h(x)$ چند جمله‌ای پالیندرومیک در $\mathbb{F}_q[x;\theta]$، با درجه‌ی t می‌باشد.

۱. اگر درجه‌ی t فرد باشد؛ آن‌گاه $h^R(x)$ یک چند جمله ای θ ـ پالیندرومیک می‌باشد.

۲. اگر درجه‌ی t زوج باشد؛ آن‌گاه $h^R(x)$ چند جمله ای پالیندرومیک می‌باشد.

برهان. فرض کنید $h(x) = h_0 + h_1 + \cdots + h_{t-1}x^{t-1} \in \mathbb{F}_q[x;\theta]$ چند جمله‌ای پالیندرومیک باشد. آن‌گاه برای هر $i = 0, 1, \ldots, t-1$ داریم $h_i = h_{t-i}$ لذا

$$h^R(x) = \sum_{i=0}^{t} x^i a_i = \sum_{i=0}^{t} \theta(h_t - 1)x^i.$$

۱. فرض کنید درجه‌ی t فرد باشد؛ در این صورت اگر i عددی فرد باشد، آن‌گاه $t-i$ زوج می‌شود. داریم: $a_i = \theta^i(h_{t-i}) = \theta(h_{t-i}) = \theta(h_i)$ و $a_{t-i} = \theta^{t-i}(h_i) = h_i$ یعنی $a_i = \theta(a_{t-i})$ جایی که i زوج باشد. می‌توان نتیجه گرفت که $h^R(x)$ یک چند جمله ای θ ـ پالیندرومیک می‌باشد.

۲. فرض کنید درجه‌ی t زوج باشد. در این صورت اگر i عددی فرد باشد، آن‌گاه $t-i$ فرد می‌باشد. داریم: $a_i = \theta^i(h_{t-i}) = \theta(h_{t-i}) = \theta(h_i)$ ، و $a_{t-i} = \theta^{t-i}(h_i) = \theta(h_i)$ یعنی $a_i = a_{t-i}$ جایی که i زوج باشد. می‌توان نتیجه گرفت $h^R(x)$ یک چندجمله‌ای پالیندرومیک می‌باشد.

□

نتیجه ۲.۴.۳. فرض کنید C یک کد دوری اریب از طول n با اتومورفیسم θ روی میدان $\mathbb{F}_q$ می‌باشد. در این صورت اگر C یک کد DNA برگشت‌پذیر باشد، آن‌گاه $C^{\perp}$ نیز یک کد DNA برگشت‌پذیر می‌باشد.

برهان. فرض کنید C کد دوری اریب باشد و $g(x)$ یک چندجمله‌ای ناصفر از کمترین درجه در C باشد. آن‌گاه C توسط چندجمله‌ای $g(x)$ تولید می‌شود و $g(x)$ یک مقسوم علیه راست از $x^n - ۱$ در حلقه‌ی $\mathbb{F}_q[x;\theta]$ می‌باشد. (مرجع [۹] رابببینید)
فرض کنید C یک کد DNA برگشت‌پذیر باشد. در این صورت دو حالت داریم:

۱. اگر درجه‌ی چندجمله‌ای $g(x)$ عدد صحیح زوج باشد، در این صورت با استفاده از قضیه ۱.۴.۳ می‌توان نتیجه گرفت $g(x)$ یک چند جمله‌ای پالیندرومیک می‌باشد. اگر $x^n - ۱ = h(x)g(x)$ در $\mathbb{F}_q[x;\theta]$، با توجه به نتیجه ۱.۴.۳ و قضیه ۵.۴.۳، می‌توان نتیجه گرفت $h^R(x)$ چندجمله‌ای پالیندرومیک می‌باشد. از این رو $C^{\perp} = \langle h^R(x)\rangle$ یک کد DNA برگشت‌پذیر است. (طبق قضیه ۱.۴.۳).

۲. اگر درجه‌ی چند جمله‌ای $g(x)$ عدد صحیح فرد باشد، در این صورت C می تواند توسط چندجمله‌ای θـ پالیندرومیک $\acute{g}(x)$ جایی که مضرب چندجمله‌ای $g(x)$ باشد، تولید شود. بنابراین $\acute{g}(x)$ یک مقسوم علیه راست از $x^n - ۱$ در $\mathbb{F}[x;\theta]$. در این صورت اگر $x^n - ۱ = \acute{g}(x)\acute{h}(x)$ در $\mathbb{F}[x;\theta]$ می‌توان نتیجه گرفت طبق قضیه ۴.۴.۳ و ۵.۴.۳ $\acute{h}^R(x)$ نیز یک چندجمله‌ای θ ـ پالیندرومیک می‌باشد. جایی که $C^{\perp} = \langle \acute{h}^R(x)\rangle$ یک کد DNA برگشت‌پذیر است. (قضیه ۲.۴.۳ ببینید).

□

مثال ۱.۴.۳. فرض کنیم α یک عنصر اولیه از میدان $\mathbb{F}_{۱۶}$، جایی که $\alpha^۴ = \alpha + ۱$. آن‌گاه $x^۶ - ۱ = h(x)g(x) = (۱ + \alpha^۷x + \alpha^۷x^۲ + x^۳)(۱ + \alpha^۷x + \alpha^{۱۳}x^۲ + x^۳)$ در حلقه‌ی $\mathbb{F}_{۱۶}[x;\theta]$ باشد. چون $g(x)$ یک چندجمله‌ای θ ـ پالیندرومیک با درجه‌ی فرد می‌باشد، لذا کد دوری اریب تولید شده توسط $C = \langle g(x)\rangle$، یک کد DNA برگشت‌پذیر با پارامترهای $[۶,۳,۴]$ روی

میدان $\mathbb{F}_{۱۶}$ می‌باشد. در این صورت $C^{\perp}$ نیز یک کد دوری برگشت‌پذیر که توسط

$$h^R(x) = ۱ + \alpha^{۱۳}x + \alpha^{۷}x^{۲} + x^{۳}$$

تولید می‌شود، چون $h^R(x)$ یک چندجمله‌ای θ ـ پالیندرومیک می‌باشد، لذا کد دوگان $C^{\perp}$ نیز یک کد DNA برگشت‌پذیر است.

مثال ۲.۴.۳. فرض کنیم α یک عنصر اولیه از میدان $\mathbb{F}_{۱۶}$، باشد جایی که $\alpha^{۴} = \alpha + ۱$. آن‌گاه $x^{۱۰} - ۱ = h(x)g(x) = (۱+\alpha x+\alpha^{۳}x^{۲}+x^{۳}+\alpha x^{۳}+x^{۴})(۱+\alpha x+\alpha^{۱۱}x^{۲}+\alpha^{۱۱}x^{۴}+\alpha x^{۵}+x^{۶})$ در حلقه‌ی $\mathbb{F}_{۱۶}[x;\theta]$ باشد. چون $g(x)$ یک چندجمله‌ای θ ـ پالیندرومیک با درجه‌ی فرد می‌باشد، لذا کد دوری اریب تولید شده توسط $C = \langle g(x)\rangle$، یک کد DNA برگشت‌پذیر از طول ۱۰ روی میدان $\mathbb{F}_{۱۶}$ می‌باشد. دوگان کد C، $C^{\perp} = \langle h^R(x)\rangle = \langle(۱ + \alpha^{۴}x + \alpha^{۳}x^{۲} + \alpha^{۴}x^{۳} + x^{۴})\rangle$ یک کد دوری اریب می‌باشد. چون $h^R(x)$ یک چندجمله‌ای پالیندرومیک از درجه‌ی زوج می‌باشد، لذا $C^{\perp}$ نیز یک کد DNA برگشت‌پذیر می‌باشد.

۵.۳ کدهای DNA با طول فرد

فرض کنیم m مرتبه‌ی اتومورفیسم θ باشد. در مرجع [۲۲]، عوامل $x^n - ۱$ در $\mathbb{F}_q[x;\theta]$، جایی که $(n,m) = ۱$، تعیین شده‌اند. برای بحث ما، چون $q = ۴^{۲s}$ و مرتبه‌ی اتومورفیسم θ برابر ۲ می باشد، لذا لم زیر یک نتیجه‌ی مستقیم از لم ۲ در مرجع [۲۲] می‌باشد.

لم ۱.۵.۳. فرض کنید $g(x)$ یک مقسوم‌علیه راست از $x^n - ۱$ در حلقه‌ی $\mathbb{F}_{۴^{۲s}}[x;\theta]$، است جایی که مرتبه θ برابر ۲ و n فرد می‌باشد. در این صورت، $g(x)$ یک چندجمله‌ای روی میدان $\mathbb{F}_{۴^s}$ می‌باشد. علاوه بر این تجزیه $x^n - ۱$ در $\mathbb{F}_{۴^{۲s}}[x;\theta]$، مشابه تجزیه $x^n - ۱$ در حلقه‌ی جابه‌جایی $\mathbb{F}_{۴^s}$ می‌باشد.

می‌توان نتیجه گرفت اگر C کد دوری اریب با طول فرد روی $\mathbb{F}_{۴^s}$ با اتومورفیسم θ باشد، آن‌گاه C یک کد دوری معمولی در $\mathbb{F}_{۴^{۲s}}$ تولید شده توسط یک چندجمله‌ای با ضرایب در $\mathbb{F}_{۴^s}$ می‌باشد. در واقع، $\mathbb{F}_{۴^s}$ زیرمیدان ثابت $\mathbb{F}_{۴^{۲s}}$ تحت θ می‌باشد، زیرا برای هر $\beta \in \mathbb{F}_{۴^s}$ داریم: $\theta(\beta) = \beta^{۴^s} = \beta$. اگر $g(x) \in \mathbb{F}_{۴^{۲s}}[x;\theta]$، یک چندجمله‌ای پالیندرومیک با ضرایب در $\mathbb{F}_{۴^s}$ باشد، آن‌گاه $g(x)$ یک چندجمله‌ای θ ـ پالیندرومیک نیز می‌باشد.

قضیه ۱.۵.۳. فرض کنید $x^n - ۱ = h(x)g(x)$ در حلقه‌ی $\mathbb{F}_{۴^{۲s}}[x;\theta]$، جایی که n عددی فرد باشد. در این صورت، کد دوری اریب $C = \langle g(x)\rangle$ از طول n روی میدان $\mathbb{F}_{۴^{۲s}}$، یک کد DNA برگشت‌پذیر است، اگر و تنها اگر $g(x)$ یک چندجمله‌ای پالیندرومیک باشد.

قضیه ۲.۵.۳. فرض کنید C یک کد دوری اریب باشد که توسط مقسوم‌علیه $x^n - ۱$ در حلقه‌ی $\mathbb{F}_{۴^{۲s}}[x;\theta]$ تولید می‌شود؛ جایی که n فرد باشد. در این صورت اگر C یک کد DNA برگشت‌پذیر باشد، آن‌گاه کد دوگان $C^{\perp}$ نیز یک کد DNA برگشت‌پذیر است.

مثال ۱.۵.۳. فرض کنیم α یک عنصر اولیه از میدان $\mathbb{F}_{۱۶}$ باشد. در این صورت، زیرمیدان ثابت تحت θ برابر است با $\mathbb{F}_۴ = \left\{۰, ۱, \alpha^۵, \alpha^{۱۰}\right\}$ و $x^۵ - ۱ = (x-۱)(x^۲ + \alpha^۵ x + ۱)(x^۲ + \alpha^{۱۰} x + ۱)$ درحلقه‌ی $\mathbb{F}_{۱۶}[x;\theta]$.

فرض کنیم $g(x) = x^۲ + \alpha^{۱۰} x + ۱$. در این صورت کد دوری اریب $C = \langle g(x)\rangle$ هم یک کد DNA برگشت‌پذیر و هم یک کد برگشت‌پذیر با پارامترهای $[۵,۳,۳]$ روی میدان $\mathbb{F}_{۱۶}$ می‌باشد. اینجا داریم

$$h(x) = h^R(x) = x^۳ + \alpha^{۱۰} x^۲ + \alpha^{۱۰} x + ۱.$$

چون $h^R(x)$ یک چندجمله‌ای پالیندرومیک می‌باشد، لذا کد دوگان $C^{\perp} = \langle h^R(x)\rangle$ نیز یک کد DNA برگشت‌پذیر می‌باشد.

نتیجه ۱.۵.۳. براساس مباحثی که در این پایان‌نامه بررسی شد می‌توان نتیجه گرفت که با وجود تفاوت آشکار میان نظریه کد و زیست‌شناسی سلولی، اما مشابهت‌هایی از لحاظ حفاظت از داده‌ها و شروع و توقف خواندن پیام وجود دارد. کدهای ژنتیکی مانند کدهای دودویی حاوی سیگنال‌هایی هستند که مکان شروع و پایان خواندن پیام را مشخص می‌کنند. کد ژنتیک یک کدون شروع و سه کدون پایانی دارد. در کدها بیت ناصحیح می‌تواند مشکلات قابل توجهی در انتقال پیام به وجود آورد. در DNA تغییرات نوکلئوتیدها منجر به تغییر ارزش و مقادیر کدون برای آمینواسید می‌شود. همانطور که اشاره شد پژوهش‌ها و الگوریتم و قضیه‌های مطرح شده خیلی با پیشنهاد درمان فاصله دارند، ولی تاثیر زیادی در شناخت بسیاری از بیماری‌ها مانند سرطان دارند.

مراجع

[1] Abualrub, Taher, Ghrayeb, Ali, Aydin, Nuh, and Siap, Irfan. On the construction of skew quasi-cyclic codes. *IEEE Transactions on Information Theory*, 56(5):2081–2090, 2010.

[2] Abualrub, Taher, Ghrayeb, Ali, and Zeng, Xiang Nian. Construction of cyclic codes over gf (4) for dna computing. *Journal of the Franklin Institute*, 343(4-5):448–457, 2006.

[3] Adleman, Leonard M. Molecular computation of solutions to combinatorial problems. *Science*, pages 1021–1024, 1994.

[4] Aghagolzadeh, Mehdi, Eldawlatly, Seif, and Oweiss, Karim. Coding stimulus information with cooperative neural populations. In *2009 IEEE International Symposium on Information Theory*, pages 1594–1598. IEEE, 2009.

[5] Anastassiou, Dimitris. Genomic signal processing. *IEEE signal processing magazine*, 18(4):8–20, 2001.

[6] Battail, Gérard. Should genetics get an information-theoretic education? *IEEE Engineering in Medicine and Biology Magazine*, 25(1):34–45, 2006.

[7] Bayram, Aysegul, Oztas, Elif Segah, and Siap, Irfan. Codes over $f_4 + vf_4$ and some dna applications. *Designs, Codes and Cryptography*, 80(2):379–393, 2016.

[8] Borodovsky, M et al. Genemark: A family of gene prediction programs.

[9] Boucher, Delphine, Geiselmann, Willi, and Ulmer, Félix. Skew-cyclic codes. *Applicable Algebra in Engineering, Communication and Computing*, 18(4):379–389, 2007.

[10] Boucher, Delphine, Geiselmann, Willi, and Ulmer, Félix. Skew-cyclic codes. *Applicable Algebra in Engineering, Communication and Computing*, 18(4):379–389, 2007.

[11] Boucher, Delphine and Ulmer, Felix. Coding with skew polynomial rings. *Journal of Symbolic Computation*, 44(12):1644–1656, 2009.

[12] Boucher, Delphine and Ulmer, Felix. Coding with skew polynomial rings. *Journal of Symbolic Computation*, 44(12):1644–1656, 2009.

[13] Boucher, Delphine and Ulmer, Felix. A note on the dual codes of module skew codes. In *IMA International Conference on Cryptography and Coding*, pages 230–243. Springer, 2011.

[14] Cridge, Andrew G, Crowe-McAuliffe, Caillan, Mathew, Suneeth F, and Tate, Warren P. Eukaryotic translational termination efficiency is influenced by the 3′ nucleotides within the ribosomal mrna channel. *Nucleic acids research*, 46(4):1927–1944, 2018.

[15] Dayan, Peter and Abbott, Laurence F. Theoretical neuroscience: computational and mathematical modeling of neural systems. 2001.

[16] Dimitrov, Alexander G and Miller, John P. Neural coding and decoding: communication channels and quantization. *Network: Computation in Neural Systems*, 12(4):441–472, 2001.

[17] Dimitrov, Alexander G, Parker, Albert E, and Gedeon, Tomas. Symmetry breaking in soft clustering decoding of neural codes. *BMC Neuroscience*, 10(1):P120, 2009.

[18] Dingel, Janis and Milenkovic, Olgica. A list-decoding approach for inferring the dynamics of gene regulatory networks. In *2008 IEEE International Symposium on Information Theory*, pages 2282–2286. IEEE, 2008.

[19] Doi, Eizaburo, Balcan, Doru C, and Lewicki, Michael S. A theoretical analysis of robust coding over noisy overcomplete channels. In *Advances in neural information processing systems*, pages 307–314, 2006.

[20] Freeland, Stephen J, Wu, Tao, and Keulmann, Nick. The case for an error minimizing standard genetic code. *Origins of Life and Evolution of the Biosphere*, 33(4-5):457–477, 2003.

[21] Grassl, M. Table of bounds on linear codes {Online}, 1995.

[22] Gursoy, Fatmanur, Siap, Irfan, and Yildiz, Bahattin. Construction of skew cyclic codes over $f_q + v f_q$. *Adv. Math. Commun*, 8(3):313–322, 2014.

[23] Hammons, A Roger, Kumar, P Vijay, Calderbank, A Robert, Sloane, Neil JA, and Solé, Patrick. The z/sub 4/-linearity of kerdock, preparata, goethals, and related codes. *IEEE Transactions on Information Theory*, 40(2):301–319, 1994.

[24] Jacobson, Nathan. The theory of rings, new york. In *Amer. Math. Soc*, volume 61, pages 875–881, 1943.

[25] Johnson, Don H and Ray, Will. Optimal stimulus coding by neural populations using rate codes. *Journal of computational neuroscience*, 16(2):129–138, 2004.

[26] Kumar, K Raj, Pakzad, Payam, Salavati, Amir Hesam, and Shokrollahi, Amin. Phase transitions for mutual information. In *2010 6th International Symposium on Turbo Codes & Iterative Information Processing*, pages 137–141. IEEE, 2010.

[27] Liebovitch, Larry S, Tao, Yi, Todorov, Angelo T, and Levine, Leo. Is there an error correcting code in the base sequence in dna? *Biophysical Journal*, 71(3):1539–1544, 1996.

[28] Liebovitch, Larry S, Tao, Yi, Todorov, Angelo T, and Levine, Leo. Is there an error correcting code in the base sequence in dna? *Biophysical Journal*, 71(3):1539–1544, 1996.

[29] Ling, San and Xing, Chaoping. *Coding theory: a first course*. Cambridge University Press, 2004.

[30] Mac Donaill, DA. The role of error-coding in shaping the nucleotide alphabet: nature's choice of a, u, c and g. In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)*, volume 4, pages 3850–3853. IEEE, 2003.

[31] Mac Dónaill, Dónall A. Why nature chose a, c, g and u/t: an error-coding perspective of nucleotide alphabet composition. *Origins of Life and Evolution of the Biosphere*, 33(4-5):433–455, 2003.

[32] May, Elebeoba E, Vouk, Mladen A, Bitzer, Donald L, and Rosnick, David I. Coding theory based models for protein translation initiation in prokaryotic organisms. *BioSystems*, 76(1-3):249–260, 2004.

[33] McDonald, Bernard R. *Finite rings with identity*, volume 28. Marcel Dekker Incorporated, 1974.

[34] McDonald, Bernard R. *Finite rings with identity*, volume 28. Marcel Dekker Incorporated, 1974.

[35] Milenkovic, Olgica and Vasic, Bane. Information theory and coding problems in genetics. In *Information Theory Workshop*, pages 60–65. IEEE, 2004.

[36] Ore, Oystein. Theory of non-commutative polynomials. *Annals of mathematics*, pages 480–508, 1933.

[37] Oztas, Elif Segah and Siap, Irfan. Lifted polynomials over ${}_16$ and their applications to dna codes. *Filomat*, 27(3):459–466, 2013.

[38] Oztas, Elif Segah and Siap, Irfan. On a generalization of lifted polynomials over finite fields and their applications to dna codes. *International Journal of Computer Mathematics*, 92(9):1976–1988, 2015.

[39] Prange, Eugene. *Cyclic error-correcting codes in two symbols*. Air force Cambridge research center, 1957.

[40] Prange, Eugene. Some cyclic error-correcting codes with simple decoding algorithms. *AFCRC-TN-58-156*, 1985.

[41] Rieke, Fred, Warland, David, Van Steveninck, Rob De Ruyter, Bialek, William S, et al. *Spikes: exploring the neural code*, volume 7. MIT press Cambridge, 1999.

[42] Rosen, G. Examining coding structure and redundancy in dna. *IEEE engineering in medicine and biology magazine*, 25(1):62–68, 2006.

[43] Rosen, Gail L. Finding near-periodic dna regions using a finite-field framework. In *Proc. IEEE Workshop Genomic Signal Processing (GENSIPS)*. Citeseer, 2004.

[44] Rosen, Gail L and Moore, Jeffrey D. Investigation of coding structure in dna. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).*, volume 2, pages II–361. IEEE, 2003.

[45] Shen, Dan and Cruz, Jose B. Encoding strategy for maximum noise tolerance bidirectional associative memory. *IEEE Transactions on Neural Networks*, 16(2):293–300, 2005.

[46] Siap, Irfan, Abualrub, Taher, Aydin, Nuh, and Seneviratne, Padmapani. Skew cyclic codes of arbitrary length. *IJICoT*, 2(1):10–20, 2011.

[47] Siap, Irfan, Abualrub, Taher, and Ghrayeb, Ali. Similarity cyclic dna codes over rings. In *2008 2nd International Conference on Bioinformatics and Biomedical Engineering*, pages 612–615. IEEE, 2008.

[48] Siap, Irfan, Abualrub, Taher, and Ghrayeb, Ali. Cyclic dna codes over the ring $f_2[u]/(u^2-1)$ based on the deletion distance. *Journal of the Franklin Institute*, 346(8):731–740, 2009.

[49] Sicot, Guillaume and Pyndiah, Ramesh. Study on the genetic code: comparison with multiplexed codes. In *2007 IEEE International Symposium on Information Theory*, pages 2666–2670. IEEE, 2007.

[50] Stamatakos, Georgios S, Dionysiou, Dimitra D, Graf, Norbert M, Sofra, Nikoletta A, Desmedt, Christine, Hoppe, Alexander, Uzunoglu, Nikolaos K, and Tsiknakis, Manolis. The" oncosimulator": a multilevel, clinically oriented simulation system of tumor growth and organism response to therapeutic schemes. towards the clinical evaluation of in silico oncology. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6628–6631. IEEE, 2007.

[51] Toyoizumi, Taro, Pfister, Jean-Pascal, Aihara, Kazuyuki, and Gerstner, Wulfram. Generalized bienenstock–cooper–munro rule for spiking neurons that maximizes information transmission. *Proceedings of the National Academy of Sciences*, 102(14):5239–5244, 2005.

[52] Yildiz, Bahattin and Siap, Irfan. Cyclic codes over $f_2[u]/(u^4-1)$ and applications to dna codes. *Computers & Mathematics with Applications*, 63(7):1169–1176, 2012.

[۵۳] هاشمی، رضا یوسفی ـ مریم. زیست‌شناسی سلولی و ملکولی. پایان‌نامه کارشناسی‌ارشد.

واژه‌نامه فارسی به انگلیسی

آبلی . . . . . Abelian

الفبای کد . . . . . Code Alphabet

ایده‌آل . . . . . Ideal

ایده‌آل چپ . . . . . Left ideal

ایده‌آل راست . . . . . Right ideal

انحراف از قالب . . . . . Framing offset

پایه متعامد . . . . . Orthogonal basis

پوچ . . . . . Null

تجزیه . . . . . Decomposition

تقسیم . . . . . Division

تقسیم سلولی . . . . . Cell division

تکرار . . . . . Replication

جهش . . . . . Mutation

جهش تغییر قالب . . . . . Frame shift mutation

جهش سرکوب‌گر . . . . . Suppressor mutation

چندجمله‌ای . . . . . Polynomial

چندجمله‌ای خود متقابل . . . . . Self- reciprocal polynomial

حلقه . . . . . Ring

حلقه چندجمله‌ای اریب . . . . . Skew polynomial ring

خودریختی . . . . . Automorphism

دامنه اقلیدسی . . . . . Euclidean domain

دامنه . . . . . Domain

درجه ورودی . . . . . in-degree

رشته . . . . . String

زیرفضا . . . . . Subspace

سلول . . . . . Cell

سیتوپلاسم . . . . . Cytoplasm

شبکه تنظیم ژن . . . . . Gen regulatory network

ضریب . . . . . Coefficient

عنصراولیه . . . . . Primitive element

کد DNA برگشت‌پذیر . . . . . Reversible DNA code

کد تصحیح خطا . . . . . Error correction code

کد ژنتیک . . . . . Genetic code

کد دوری اریب . . . . . Skew cyclic code

کدگذاری . . . . . Coding

کدگذاری کانال . . . . . Channel coding

کدگذاری منبع . . . . . Source coding

واژه‌نامه انگلیسی به فارسی

آبلی . . . . . Abelian

خودریختی . . . . . Automorphism

سلول . . . . . Cell

تقسیم سلولی . . . . . Cell division

کدگذاری کانال . . . . . Channel coding

کدگذاری . . . . . Coding

الفبای کد . . . . . Code Alphabet

ضریب . . . . . Coefficient

سیتوپلاسم . . . . . Cytoplasm

تجزیه . . . . . Decomposition

تقسیم . . . . . Division

دامنه . . . . . Domain

کدتصحیح خطا . . . . . Error Correction Code

دامنه اقلیدسی . . . . . Euclidean domain

جهش تغییر قالب . . . . . Frame shift mutation

انحراف از قالب . . . . . Framing offset

کد ژنتیک . . . . . Genetic code

شبکه تنظیم ژن . . . . . Gen regulatory network

ایده آل . . . . . Ideal

درجه ورودی . . . . . In-degree

ایده ال چپ . . . . . Left ideal

جهش . . . . . Mutation

پوچ ..... Null

پایه متعامد ..... Orthogonal basis

چندجمله‌ای ..... Polynomial

عنصر اولیه ..... Primitive element

کد DNA برگشت‌پذیر ..... Reversible DNA code

ایده ال راست ..... Right ideal

حلقه ..... Ring

چندجمله ای خودمتقابل ..... Self- reciprocal polynomial

کد دوری اریب ..... Skew cyclic code

کدگذاری منبع ..... Source coding

رشته ..... String

زیرفضا ..... Subspace

تفریق ..... Subtraction

جهش سرکوب‌گر ..... Suppressor mutation

Abstract

Today the use of mathematical sciences in biology, has a special place and is considered as a scientific benchmark for research. Presenting mathematical theories, statistics and probability, theoretical study of diseases due to disorders in the metabolic pathway or genetic changes and building computational algorithm and modeling to answer the questions of biological sciences, are some of the most significant applications of mathematics in biology. In this thesis we intend to discuss some applications of mathematics in biology, especially applications of coding theory in cellular and molecular biology. First, we study skew cyclic codes; linear codes over finite fields obtained by left ideals of the quotient ring of noncommutative skew polynomial rings.

Then, we construct reversible DNA codes by using the skew cyclic codes over $F_{4^{2s}}$. We also examine how the existence and properties of skew cyclic codes relate to the structure of reversible DNA codes.

Finally, we will show that the dual codes of skew cyclic codes are also reversible DNA codes and skew self-reciprocal polynomials are θ-palindromic.

Keywords: Genetic code; Nucleotide; Finite field; Channel coding; Decoding rule; Cyclic code; Reversible DNA codes; θ-palindromic polynomials; Skew self-reciprocal polynomial.

Faculty Of Mathematical Sciences

MSc Thesis in Applied Mathematics

Coding of Biological Systems and Gene Regulation Networks

By: Elham Bahramipour

Supervisors:

Dr. Abdollah Alhevaz
Dr. Meysam Alishahi

Advisor:

Prof. Ebrahim Hashemi

September 2020