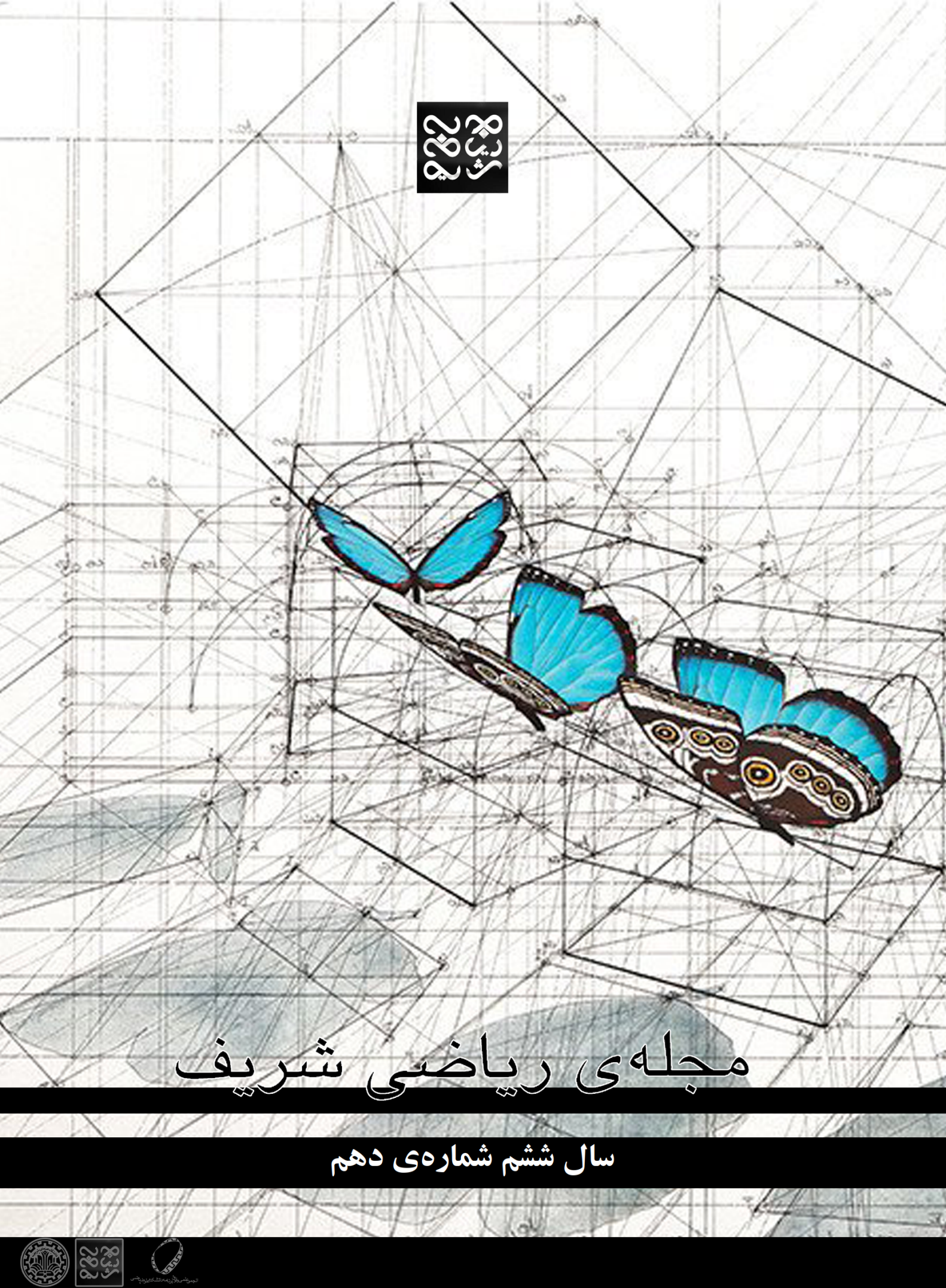


مجله‌ی ریاضی شریف

سال ششم شماره‌ی دهم

مجله ریاضی شریف

مهر ۹۷، شماره دهم

صاحب امتیاز: انجمن علمی و فوق برنامه‌ی دانشکده‌ی علوم ریاضی
مدیر مسئول: دکتر امیر جعفری
سردبیر: کیمیا کاظمیان
همکاران این شماره: امیرحسین اخلاصی، نیما افشار، کامیار امینی، تینا ترکمان، علیرضا توکلی، علی چراغی، سایه خانی‌ها، مائده خسروی‌زاده، ریحانه صابرآملی، یاشار طالبی‌راد، عطا طهوری، امیر عزیزی‌جیرآبادی، شایان غلامی، محمدعلی کرمی، هلیا محمدی‌دوستدار، سهیل معماریان، بهراد معینی، امیرحسین مکاره‌چی، حسین نادری، سیدسروش هاشمی، محمد هنری، علی یزدانی

فهرست مطالب

۱ ... قضیه آخر فرما
۱۰ ... مقدمه‌ای بر شبکه‌های عصبی (قسمت اول)
۱۸ ... تبدیل فوریه در گروه‌ها
۳۵ ... قدم زن تصادفی ساده و شبکه‌های الکتریکی
۴۰ ... سه آزمون اول بودن با پیچیدگی زمانی چندجمله‌ای
۴۷ ... همبستگی ماکسیمال
۵۴ ... دو فرهنگ ریاضی
۶۵ ... چند مساله

تحت حمایت و نظارت معاونت فرهنگی
دانشگاه صنعتی شریف

مجله‌ی ریاضی شریف
مهر ۹۷، شماره‌ی دهم

قضیه آخر فرما

علی چراغی

چکیده

در این مقاله سعی می‌شود در مورد پیشنیازها و روند اثبات قضیه آخر فرما [1] توضیح داده شود. تا آنجا که ممکن بوده، سعی شده خواندن این مقاله نیازی به هیچ پیشنیازی نداشته باشد، ولی با این وجود توضیح بعضی از پیشنیازها (کمی نظریه رسته‌ها و نظریه جبری اعداد) فضای بسیاری را اشغال می‌کند. خواننده برای یادگیری آن‌ها می‌تواند به [۶] و [۵] رجوع کند.

۱ مقدمه

در سال ۱۶۳۷، در ویرایشی از کتاب حساب [2] فرما ادعای زیر را نوشت (که آن را ترجمه کرده‌ایم!): "غیرممکن است که یک مکعب را به دو مکعب تجزیه کرد، یا یک توان چهارم را به دو توان چهارم، و در حالت کلی هر توان بزرگ‌تر از دو را به دو توان شبیه هم، من یک اثبات شگفت‌انگیز از این پیدا کرده‌ام که این حاشیه کوچک‌تر از آن است که آن را قرار دهم." پس از او ریاضیدان‌های بسیاری برای اثبات این موضوع تلاش کردند و به نتایج جزئی نیز رسیدند. مثلاً:

۱) فرما در سال ۱۶۴۰، حالت توان چهارم را حل کرد.

۲) اویلر [3] بین سال‌های ۱۷۵۸ و ۱۷۷۰ حالت توان سوم را حل کرد.

۳) سوفی ژرمن [4] در اوایل قرن ۱۹، راه‌حلی ایجاد کرد که حداقل برای همه توان‌های اول فرد کمتر از ۱۰۰، حالت اول قضیه فرما را ثابت می‌کرد ($x^n + y^n = z^n \Rightarrow n \mid xyz$).

۴) کومر [5] با استفاده از نظریه اعداد جبری قضیه را برای حدود 60 درصد (حدسی!) از اعداد اول ثابت کرد.

۵) در سال ۱۹۷۷، ترجانیان [6] حالت اول قضیه را برای همه‌ی توان‌های زوج (غیر ۲!) اثبات کرد.

ولی در سال ۱۹۸۴ بود که فری [7] یک ارتباط بین این حدس و خمی بیضوی یافت و احساس کرد که غلط بودن قضیه آخر فرما باعث رد حدسی معروف از شیمورا [8] و تانیاما [9] می‌شود. دو سال پس از آن، ریبت [10] این روند را ادامه داد و با اثبات حدسی از سر

[1] P. Fermat
[2] Arithmetica
[3] L. Euler
[4] S. Germain
[5] E. Kummer
[6] G. Terjanian
[7] G. Frey
[8] G. Shimura
[9] Y. Taniyama
[10] K. Ribet
[11] J.P. Serre
[12] Epsilon conjecture

[11]، به نام حدس اپسیلون [12]، اثبات قضیه آخر فرما را به حدس شیمورا–تانیاما کاهش داد. بعد از او، ریاضیدان انگلیسی، اندرو وایلز [13]، که از زمان کودکی آرزوی حل این مسئله را داشت، ۶ سال روی این مسئله کار کرد و در نهایت، در سال ۱۹۹۳ موفق به اثبات مقداری از حدس شیمورا–تانیاما شد که برای اثبات قضیه آخر فرما کافی بود. پس از آن اشکالی در اثبات وایلز پیدا شد که باعث شد وایلز و تیلور [14] (شاگرد قدیمی او) یک سال دیگر تلاش کنند تا آن اشکال را برطرف نمایند. در انتها، در سال ۱۹۹۵، قضیه آخر فرما به‌طور کامل اثبات شد. همچنین با قوی‌تر کردن ایده‌های وایلز، تعدادی ریاضیدان دیگر، بروی [15] و کنراد [16] و دایموند [17] و تیلور، توانستند حدس شیمورا–تانیاما را به‌طور کامل اثبات کنند و پس از آن، نام آن تبدیل به قضیه مدولاریتی شد.

در این مقاله ابتدا در مورد خم‌های بیضوی و فرم‌های مدولار کمی توضیح داده می‌شود و سپس روند اثبات قضیه آخر فرما توضیح داده می‌شود. اثبات‌ها به دلیل طولانی بودن، معمولاً ارجاع داده شده‌اند.

۲ صورت قضیه

قضیه آخر فرما، قضیه‌ای با صورت ساده در نظریه اعداد است و از این قرار است:

قضیه ۱. *فرض کنید $n \geq 3$ عددی صحیح باشد. اگر برای $A, B \in \mathbb{Z}$ داشته باشیم $A^n + B^n = C^n$ آنگاه $ABC = 0$.*

حل این مسئله بیش از ۳۵۰ سال زمان برده و با روشی کاملاً غیر مستقیم اثبات شده است.

اولاً اگر $d|n$ آنگاه داریم:

$$A^n + B^n = C^n \Leftrightarrow (A^{\frac{n}{d}})^d + (B^{\frac{n}{d}})^d = (C^{\frac{n}{d}})^d$$

پس این نشان می‌دهد که کافی است مسئله را برای n های اول فرد و $n = 4$ حل کنیم تا برای همه n ها حل شود. حالت $n = 4$ را خود فرما با روش نزول نامتناهی خود ثابت کرده است. پس کافیست قضیه آخر فرما را برای n های اول فرد حل کنیم. همان‌گونه که در مقدمه گفته شد، این قضیه برای اعداد اول بسیاری قبل از اثبات وایلز حل شده بود ولی حالت کلی آن باقی مانده بود.

پس ما از این به بعد یک عدد اول فرد p را فیکس کرده و فرض می‌کنیم $A^p + B^p = C^p$ برای $ABC \neq 0$ به‌طوری که $\gcd(A, B, C) = 1$ و سعی می‌کنیم تناقضی یافت کنیم!

۳ خم‌های بیضوی

منظور از یک خم بیضوی روی اعداد گویا، یک خم جبری هموار [18] است به‌طوری که با معادله‌ی زیر داده شده باشد:

$$y^2 = x^3 + ax + b, \quad a, b \in \mathbb{Q}$$

در این صورت هموار بودن به معنی این خواهد بود که چندجمله‌ای $x^3 + ax + b$ ریشه مکرر نداشته باشد (در حالت کلی هموار بودن به این معنی است که در هر نقطه از خم، یک خط مماس بر خم موجود باشد). منظور از تفکیک‌کننده [19] برای این خم عدد زیر است: $\Delta = -16(4a^3 + 27b^2)$.

تفکیک‌کننده این خاصیت را دارد که ناصفر است اگر و تنها اگر خم هموار باشد. فری خم زیر را برای بررسی پیشنهاد کرد (زیرا این خم دارای تفکیک‌کننده‌ای است که به نظر متناقض با حدس‌های اخیر بوده است):

$$F : y^2 = x(x - A^p)(x + B^p)$$

در واقع اگر تفکیک‌کننده مینیمال (مینیمم قدرمطلق تفکیک‌کننده‌ی خم‌های ایزومورف با این خم) این خم بیضوی (همواری از

[13] A. Wiles
[14] R. Taylor
[15] C. Breuil
[16] B. Conrad
[17] F. Diamond
[18] Smooth
[19] Discriminant

$ABC \neq 0$ بدست آمده است) را حساب کنیم، داریم:

$$\Delta_F = 2^{-8}(ABC)^p$$

این توان pـام کامل داشتن، یکی از موارد بنظر متناقض با آن حدس اخیر بوده است! آن حدس را بعدتر بیان می‌کنیم.

در اینجا عمل گروه روی خم بیضوی را بیان می‌کنیم. برای این کار یک خم بیضوی E با نقطه‌ی گویای $O = [0:1:0]$ فیکس کنید (نقاط پروژکتیو آن را در نظر بگیرید که برابر با اجتماع نقاط آفین آن و O است). می‌خواهیم روی $E(\mathbb{Q})$ یک عمل گروه تعریف کنیم. پس دو نقطه $P, Q \in E(\mathbb{Q})$ را در نظر می‌گیریم و جمع آن‌ها را به شکل هندسی نشان داده شده در شکل ۱ تعریف می‌کنیم. همچنین اگر بخواهیم نقطه‌ای را با خودش جمع کنیم، خط گذرنده از آن نقطه و خودش (خط مماس بر خم در آن نقطه) را در نظر می‌گیریم. این عمل خوش‌تعریف است، زیرا می‌توان به سادگی از معادله‌ی خطوط و خم نتیجه گرفت $P + Q \in E(\mathbb{Q})$.

یک خم بیضوی E تعریف شده روی اعداد صحیح $(a, b \in \mathbb{Z})$ را می‌توان به پیمانه یک عدد اول l در نظر گرفت و اگر $l \nmid \Delta_E$، خم بیضوی جدیدی را پیدا کرد. در واقع خم بیضوی جدید $\bar{E} : y^2 = x^3 + \bar{a}x + \bar{b}$ که $(\bar{.}) : \mathbb{Z} \to \mathbb{Z}/l\mathbb{Z} \cong \mathbb{F}_l$ تابع تصویر است، را برابر با کاهش [۲۰] خم بیضوی به پیمانه l تعریف می‌کنیم. حال اگر $l \mid \Delta$، مشکل این خواهد بود که خم کاهش‌یافته در دقیقاً یک نقطه ناهموار خواهد بود و پس دو حالت داریم:

۱) خم در آن نقطه ناهمواری، دوخط مماس متفاوت دارد که در این صورت به آن کاهش ضربی [۲۱] گوییم.

۲) خم در آن نقطه ناهمواری، یک خط مماس (مکرر) دارد که در این صورت به آن کاهش جمعی [۲۲] گوییم.

اصطلاحات "ضربی" و "جمعی" تفسیر ساده‌ای دارند که البته اهمیتی در بحث ما ندارند و پس آن را بیان نمی‌کنیم.

حال خم فری F، این خاصیت را دارد که در همه‌ی کاهش‌های ناهموار، کاهش ضربی دارد. در واقع فرض کنید داشته باشیم $l \mid 2^{-8}(ABC)^p$. پس مثلاً $l \mid A$ و پس داریم:

$$\bar{F} : y^2 = x(x)(x + \bar{B}^p)$$

این خم در نقطه‌ی $(0,0)$ ناهموار است و دقیقاً زمانی در l کاهش جمعی دارد که $\bar{B}^p = 0$، اما این یعنی $l \mid B$ و پس طبق $A^p + B^p = C^p$ داریم $l \mid C$، پس $l \mid \gcd(A, B, C)$ که متناقض با فرض است. پس خم فری در هر نقطه کاهش نیمه‌پایدار دارد (کاهش خوب (که کاهش این خم، هموار باشد) یا کاهش ضربی).

تعداد نقاط خم‌های بیضوی روی میدان‌های متناهی (همان کاهش آن‌ها به پیمانه اعداد اول) از اهمیت فوق‌العاده‌ای برخوردار است. ما اعداد $a_l = l + 1 - \sharp(\bar{E}(\mathbb{F}_l))$ را در نظر می‌گیریم. در این صورت داریم $|a_l| \leq 2\sqrt{l}$ (قضیه هسه [۲۳]) و می‌توان با استفاده از این اعداد، Lـتوابع [۲۴] خم‌های بیضوی را تعریف کرد (فرض می‌کنیم خم نیمه‌پایدار باشد):

$$L(E, s) = \prod_{l \mid \Delta} \frac{1}{1 \pm l^{-s}} \prod_{l \nmid \Delta} \frac{1}{1 - a_l l^{-s} + l^{1-2s}}$$

علامت $\pm$ بستگی به یک شرط موضعی دارد (شیب‌های مماس‌های خم بیضوی کاهش یافته در $\mathbb{F}_l$ هستند یا در $\mathbb{F}_{l^2}$) طبق قضیه هسه این تابع برای $Re\, s > \frac{3}{2}$ همگرا است (چرا؟).

۴ فرم‌های مدولار

منظور از یک فرم مدولار روی $SL_2(\mathbb{Z})$ با وزن k، یک تابع $f : \mathbb{H} \to \mathbb{C}$ است ($\mathbb{H}$ نیم‌صفحه‌ی بالای $\mathbb{C}$ است)، به‌طوری که

$$f(\frac{az+b}{cz+d}) = (cz+d)^k f(z), z \in \mathbb{H}, \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL_2(\mathbb{Z}) \quad (۱)$$

[20] Reduction
[21] Multiplicative reduction
[22] Additive reduction
[23] H. Hasse
[24] L-functions
[25] Fourier expansion

و روی همه‌ی نقاط $\mathbb{H}$ و همچنین در "∞" تحلیلی باشد. شرط آخر به این معنی است که بسط فوریه f [۲۵] (چون داریم $f(z+1) = f(z)$ از رابطه بالا) در بی‌نهایت از صفر شروع شود:

$$f(q) = \sum_{n\geq 0} a_n q^n, \quad q = e^{2\pi i z},\ a_n \in \mathbb{C}$$

همه‌ی فرم‌های مدولار از وزن k روی $SL_2(\mathbb{Z})$ را با $M_k(SL_2(\mathbb{Z}))$ نمایش می‌دهیم. همچنین واضح است که این تعریف را می‌توان زیرگروه‌هایی از $SL_2(\mathbb{Z})$ مثل Γ تعمیم داد به این شکل که در شرط (۱) به جای $\begin{bmatrix} a\ b \\ c\ d \end{bmatrix} \in SL_2(\mathbb{Z})$ داشته باشیم $\begin{bmatrix} a\ b \\ c\ d \end{bmatrix} \in \Gamma$. فرض می‌کنیم Γ خیلی کوچک نباشد یعنی $\Gamma(N) \subseteq \Gamma$ برای Nی که

$$\Gamma(N) = \{ \begin{bmatrix} a\ b \\ c\ d \end{bmatrix} \in SL_2(\mathbb{Z}) \mid \begin{bmatrix} a\ b \\ c\ d \end{bmatrix} \equiv \begin{bmatrix} 1\ 0 \\ 0\ 1 \end{bmatrix} \mod N\}$$

هم‌چنین قرار می‌دهیم:

$$\Gamma_0(N) = \{ \begin{bmatrix} a\ b \\ c\ d \end{bmatrix} \in SL_2(\mathbb{Z}) \mid c \equiv 0 \mod N\}$$

در این صورت $\Gamma(N)$ از اندیس متناهی در $SL_2(\mathbb{Z})$ است (چرا؟) و پس Γ و $\Gamma_0(N)$ نیز از اندیس متناهی هستند. به چنین زیرگروه‌هایی از $SL_2(\mathbb{Z})$ زیرگروه‌های هم‌نهشتی [۲۶] گوییم. پس مانند قبل، فرم‌های مدولار در این فضا را با $M_k(\Gamma)$ نمایش می‌دهیم. فرم‌های مدولار این خاصیت مهم را دارند که هر کدام از $M_k(\Gamma)$ها متناهی‌بعد هستند (پس مثلاً اگر یک فرم مدولار داشته باشیم، می‌توانیم آن را برحسب پایه‌ای در این فضا بنویسیم و ضرایب فوریه آن فرم مدولار را بهتر بشناسیم).

منظور از یک فرم کاسپ [۲۷] از وزن k روی Γ، یک فرم مدولار با وزن k روی Γ است به‌طوری که در همه‌ی "کاسپ"های Γ صفر شود: اگر Γ یک زیرگروه هم‌نهشتی باشد، می‌توان فضای $\mathbb{H}/\Gamma$ را در نظر گرفت. در این صورت این یک رویه ریمانی باز است بطوری که فشرده‌سازی [۲۸] آن به متناهی نقطه نیاز دارد. به این متناهی نقطه، کاسپ‌های Γ گوییم. فضای همه‌ی فرم‌های کاسپ از وزن k روی Γ را با $S_k(\Gamma)$ نمایش می‌دهیم. همچنین منظور از یک فرم مدولار (فرم کاسپ) از وزن k و مرحله‌ی n، یک فرم مدولار (فرم کاسپ) در $M_k(\Gamma_0(N))$ $(S_k(\Gamma_0(N)))$ است.

حال عملگرهای هکه [۲۹] را مختصراً توضیح می‌دهیم. برای سادگی فرض کنید $\Gamma = SL_2(\mathbb{Z})$. اولاً دقت کنید که هر کدام از نقاط بالای صفحه را می‌توان با یک شبکه در $\mathbb{C}$ نمایش داد:

$$\mathbb{H} \to \mathcal{L}$$

$$\alpha \mapsto < \alpha, 1 >$$

که $\mathcal{L}$، مجموعه‌ی همه‌ی شبکه‌هاست و همچنین اگر شبکه‌ها را در حد هموتتی [۳۰] (ضرب در یک عدد مختلط) در نظر بگیریم، تابع بالا یک دوسویی خواهد داد:

$$\mathbb{H} \to \mathcal{L}/\text{homothety}$$

پس فرم‌های مدولار را می‌توان روی شبکه‌ها نیز تعریف کرد. پس یک فرم مدولار از وزن k مثل f فیکس کنید. در این صورت قرار می‌دهیم:

$$f(< \omega_1, \omega_2 >) = (\omega_2)^{-k} f(\omega_1/\omega_2)$$

حال مثلاً یک شبکه L در نظر بگیرید و جمع صوری زیر را در نظر بگیرید $(n \in \mathbb{N})$:

$$T_n L := \sum_{[L:L']=n} L'$$

و برای $f \in M_k(SL_2(\mathbb{Z}))$ قرار دهید:

$$T_n f(L) = n^{k-1} \sum_{[L:L']=n} f(L')$$

در این صورت T_nها عملگرهایی روی $M_k(SL_2(\mathbb{Z}))$ و $S_k(SL_2(\mathbb{Z}))$ خواهند بود. همچنین برای $\lambda \in \mathbb{C}^*$، $[\lambda]$ را برابر با $L \mapsto \lambda L$ تعریف کنید و برای $f \in M_k(SL_2(\mathbb{Z}))$ قرار دهید:

$$([\lambda]f)(L) = f(\lambda L)$$

[26] Congruence subgroups
[27] Cusp form
[28] Compactification
[29] E. Hecke
[30] Homothety

$[\lambda]$ نیز عملگری از فضاهای فرم‌های مدولار و فرم‌های کاسپ خواهد بود.

حال می‌توان به‌سادگی چک کرد که T_n ها و $[\lambda]$ ها با یکدیگر جابجا می‌شوند و T_nها نسبت به ضرب داخلی پترسون:

$$(f,g) := \int_{\mathbb{H}/SL_2(\mathbb{Z})} f(\tau)\bar{g}(\tau)(\mathrm{Im}\,\tau)^k d\nu(\tau),$$
$$\nu(\tau) = y^{-2}dxdy, f,g \in S_k(SL_2(\mathbb{Z}))$$

خودالحاق هستند. پس می‌توان فضای فرم‌های کاسپ از وزن k روی $SL_2(\mathbb{Z})$ را به فرم‌های ویژه همه‌ی T_nها تجزیه کرد (که یکه و عمود هستند). این فرم‌های ویژه را می‌توان صریحاً پیدا کرد و برابر با سری‌های آیزنشتاین [۳۱] می‌شوند. خواص مهمی که عملگرهای هکه دارند عبارتند از:

۱) $[\lambda][\mu] = [\mu][\lambda]$ برای $\lambda, \mu \in \mathbb{C}^*$.

۲) $[\lambda]T_n = T_n[\lambda]$ برای $\lambda \in \mathbb{C}^*$ و $n \in \mathbb{N}$.

۳) $T_nT_m = T_{nm}$ اگر m,n نسبت به هم اول باشند.

۴) $T_{l^n}T_l = T_{l^{n+1}} + lT_{l^{n-1}}[l]$ برای l اول و $n \in \mathbb{N}$.

۵) $a_1(T_nf) = a_n(f)$ که منظور از $a_n(f)$ ضریب n ـام فوریه‌ی f در بی‌نهایت است.

همچنین در حالت کلی‌تر، می‌توان این عملگرهای هکه را تعریف کرد که توابعی خطی از $M_k(\Gamma_1)$ $(S_k(\Gamma_1))$ به $M_k(\Gamma_2)$ $(S_k(\Gamma_2))$ خواهند بود. چیزی که برای ما مهم است، فقط عملگرهای هکه روی $S_k(\Gamma_0(N))$ هست که تعریف آن روشن‌کننده نیست و تعمیمی از حالت قبل خواهد بود (می‌توان به سادگی روی بسط‌های فوریه آن‌ها را تعریف کرد). همچنین خواصی که این عملگرهای هکه کلی‌تر روی $S_k(\Gamma_0(N))$ دارند، مشابه خواص بالا است (برای $l \mid N$ باید حواسمان را بیشتر جمع کنیم).

مثلاً یکی از کاربردهای این عملگرها قسمت‌هایی از حدس رامانوجان [۳۲] بوده است. در واقع تابع

$$\Delta(q) = q(1-q^n)^{24},\ q = e^{2\pi iz}$$

بسط فوریه‌ی $\eta(q) = \sum_{n\geq 1}\tau(n)q^n$ را دارد که $\tau(n)$ تابع رامانوجان است. رامانوجان حدس زده بود که

۱) تابع τ ضربی است.

۲) برای l اول و $n \in \mathbb{N}$ داریم:

$$\tau(l^{n+1}) = \tau(l)\tau(l^n) - l^{11}\tau(l^{n-1})$$

۳) $|\tau(l)| \leq 2l^{11/2}$.

حال می‌توان دید که فضای $S_{12}(SL_2(\mathbb{Z}))$ یک بعدی است و $S_{12}(SL_2(\mathbb{Z})) =< \Delta >$. پس η باید بردار ویژه‌ی همه‌ی T_nها باشد. پس طبق خاصیت ۵ باید داشته باشیم:

$$a_1(T_n\Delta) = a_n(\Delta) = \tau(n)$$

و از آنجا که $a_1(\Delta) = 1$، پس باید مقدارویژه‌ی متناظر آن $\tau(n)$ باشد و پس از خاصیت ۴ با عمل کردن به روی Δ قسمت دوم حدس بدست می‌آید. همچنین از خاصیت ۳ نیز با همین روند قسمت اول حدس بدست می‌آید. قسمت سوم حدس سخت‌تر است و از اثبات حدس‌های وی [۳۳] بدست می‌آید.

یک فرم کاسپ $f \in S_k(\Gamma_0(N))$ در نظر بگیرید به‌طوری که فرم ویژه‌ی همه‌ی T_nها $(n \nmid N)$ باشد. در این صورت خواص ۳ و ۴ عملگرهای هکه پیشنهاد می‌کنند که ما L ـ تابع ضرایب بسط فوریه‌ی آن را در نظر بگیریم:

اگر $f = \sum_{n\geq 1} a_nq^n$ قرار دهید:

$$L(f,s) = \sum_{p\nmid N} \frac{a_n}{n^s}$$

و پس از ۳ و ۴ و بسط تیلور داریم که این تابع حاصلضرب اویلری زیر را دارد:

$$L(f,s) := \prod_{p\nmid N} \frac{1}{1 - a_pp^{-s} + p^{1-2s}}$$

این حاصلضرب اویلری ما را یاد L ـ تابع خم‌های بیضوی می‌اندازد. پس احتمالاً رابطه‌ای بین فرم‌های ویژه و خم‌های بیضوی وجود دارد.

[31] Eisenstein series
[32] S. Ramanujan
[33] Weil conjectures

حال می‌توان یک بیان ساده از قضیه مدولاریتی (همان "حدس اخیر" که وعده داده بودیم!) را گفت:

قضیه ۲. *فرض کنید $E/\mathbb{Q}$ یک خم بیضوی باشد. فرض کنید l یک عدد اول باشد. در این صورت قرار دهید $a_l = l+1-\bar{E}(\mathbb{F}_l)$ (برای pهای با کاهش خوب) در این صورت $f \in S_2(\Gamma_0(N))$ وجود دارد که نرمال ($a_1(f)=1$) و "جدید" و فرم ویژه‌ی همه‌ی عملگرهای هکه باشد و $a_l(f)=a_l$ برای همه بجز متناهی l.*

"جدید" یعنی این‌که f از $S_2(\Gamma_0(d))$ $(d \mid N)$ نیامده باشد: در واقع، اگر $g \in S_2(\Gamma_0(d))$ باشد، آنگاه $h : z \mapsto g(\frac{N}{d}z)$ یک فرم کاسپ در $S_2(\Gamma_0(N))$ هست و منظور از جدید یعنی f ترکیب خطی تعدادی از این خم‌های $S_2(\Gamma_0(d))$ برای dهای کمتر نباشد.

با فرض این قضیه خواهیم داشت:

$$L(E,s) = L(f,s)$$

(روی متناهی عدد اول کنار گذاشته شده می‌توان این توابع را جوری تعریف کرد که رابطه بالا درست باشد) و پس خواص خوب (معادله تابعی و گسترش تحلیلی و ...) که برای Lـتوابع فرم‌های کاسپ ساده است به Lـتوابع خم‌های بیضوی می‌رسد.

۵ نمایش‌های گالوا

در این‌جا نمایش‌های گالوا روی خم‌های بیضوی و فرم‌های مدولار را توضیح می‌دهیم (مساوی بودن نمایش‌های گالوا نیز صورتی دیگر از قضیه مدولاریتی خواهد بود). منظور از یک نمایش گالوا یک هومومورفیسم $\rho : G_{\mathbb{Q}} \to GL_n(A)$ است که $G_{\mathbb{Q}} = \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ و A یک حلقه‌ی جابجایی و یکدار است. یک مثال از نمایش‌های گالوا نمایش دایره‌بر [۳۴] است. این نمایش یک‌بعدی به شکل زیر تعریف می‌شود: یک عدد اول l را فیکس کنید و قرار دهید

$$\chi_l : G_{\mathbb{Q}} \to \mathbb{Z}_l^*$$

به‌طوری که

$$\sigma : \zeta_{l^n} \mapsto \zeta_{l^n}^{\chi_l(\sigma)} \; n \in \mathbb{N}$$

همچنین، نمایش‌های شاخه‌ای [۳۵] و غیرشاخه‌ای [۳۶] به معنی زیر هستند:

تعریف ۱. *به نمایش ρ در عدد اول l غیرشاخه‌ای گویند هرگاه $\rho(I_l)=\{1\}$ که I_l یک گروه سکون [۳۷] در l است و در غیر این صورت به آن شاخه‌ای در l گویند.*

نمایش‌های گالوا روی خم‌های بیضوی.

یک خم بیضوی $E/\mathbb{Q}$ را در نظر بگیرید. دیدیم که عمل گروه روی E داریم و پس می‌توان نقاط تابی [۳۸] روی آن را در نظر گرفت:

$$E[n] = \{p \in E(\bar{Q}) \mid [n]p = O\}$$

در این صورت داریم:

$$E[n] \cong \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$$

حال مدول تیت [۳۹] را به شکل زیر تعریف می‌کنیم:

$$T_l(E) := \lim_{\leftarrow} E[l^n] \cong \mathbb{Z}_l \times \mathbb{Z}_l$$

که $\mathbb{Z}_l$ گروه اعداد lـتایی [۴۰] هاست.

با استفاده از مدول تیت می‌توان یک نمایش گالوا معرفی کرد، در واقع برای هر نقطه‌ی $P=(x,y) \in E[l^n]$ می‌توان $\sigma \in G_{\mathbb{Q}}$ را روی آن اثر داد و $P^\sigma = (\sigma(x),\sigma(y))$ را بدست آورد که از آنجا که عمل گروه، به شکل توابع گویا تعریف شده است، P^σ باید در $E[l^n]$ باشد. پس می‌توان نمایشی به شکل $\rho : G_{\mathbb{Q}} \to GL(E[l^n]) = GL_2(\mathbb{Z}/l^n\mathbb{Z})$ تعریف کرد که این‌ها

[34] Cyclotomic
[35] Ramified
[36] Unramified
[37] Inertia group
[38] Torsion points
[39] Tate module
[40] l-adic numbers

با سیستم وارون سازگارند و پس نمایش گالوا روی مدول تیت پیدا می‌کنیم:

$$\rho_{E,l} : G_{\mathbb{Q}} \to GL(T_l(E)) \cong GL_2(\mathbb{Z}_l)$$

نمایش‌های گالوا روی فرم‌های مدولار.

تعریف این کمی دشوارتر از حالت قبل است و قسمت اصلی ساخت آن را ارجاع می‌دهیم. در واقع برای ساخت آن ابتدا باید یک خم بیضوی (واریته آبلی در حالت کلی) بسازیم و سپس نمایش گالوا روی آن را مساوی با نمایش این فرم مدولار تعریف کنیم. پس فرض کنید $f \in S_2(\Gamma_0(N))$ طوری باشد که همه‌ی ضرایب بسط فوریه‌ی آن گویا هستند. در این صورت با استفاده از آن می‌توان یک خم بیضوی E_f ساخت به‌طوری که مدولار باشد (به معنی صورت قضیه مدولاریتی بالا) و سپس قرار می‌دهیم:

$$\rho_{f,l} := \rho_{E_f,l}$$

برای دیدن نحوه ساخت آن خم بیضوی می‌توانید به

بخش ۵.۲ از [۲] رجوع کنید.

حال صورت دومی از قضیه مدولاریتی را بیان می‌کنیم:

قضیه ۳. *فرض کنید E یک خم بیضوی روی $\mathbb{Q}$ باشد. در این صورت $f \in S_2(\Gamma_0(N))$ (که f جدید و فرم‌ویژه‌ی همه‌ی عملگرهای هکه و دارای ضرایب فوریه گویا) وجود دارد که برای همه lهای اول:*

$$\rho_{f,l} \sim \rho_{E,l}$$

علامت $\sim$ به معنی مزدوج است.

روندی که وایلز برای اثبات قضیه آخر فرما در پیش گرفت همین صورت از قضیه مدولاریتی بود و کاری که او کرد این بود که قضیه مدولاریتی را برای همه‌ی خم‌های بیضوی نیمه‌پایدار (از جمله خم فری) اثبات کرد که برای اثبات قضیه آخر فرما کافی بود.

۶ حدس اپسیلون (قضیه ریبت)

این حدس را اولین بار سر مطرح کرد و ریبت آن را اثبات کرد و اثبات قضیه آخر فرما را به حدس شیمورا–تانیاما کاهش داد. این قضیه در مورد کاهش دادن مرحله‌ی یک فرم کاسپ حرف می‌زند:

قضیه ۴. (*قضیه ریبت*) *فرض کنید q عدد اولی باشد و f در $S_2(\Gamma_0(lN))$ یک فرم کاسپ جدید نرمال $(a_1(f) = 1)$ و فرم‌ویژه همه‌ی عملگرهای هکه باشد به‌طوری که دارای نمایش‌های "مطلقاً تحویل‌ناپذیر" [۴۱] $\rho_{f,q}$ باشد و این نمایش در l غیرشاخه‌ای ("متناهی" و "صاف" [۴۲]) باشد اگر $l \neq q$ $(l = q)$. آن‌گاه $g \in S_2(\Gamma_0(N))$ نرمال و جدید وجود دارد که داریم:*

$$\rho_{f,q} \sim \rho_{g,q}$$

اثبات. رجوع شود به [۳].

متناهی و صاف معانی جزئی دارند که توضیح آنها از سطح این مقاله بالاتر است. مطلقاً تحویل‌ناپذیر یعنی به عنوان نمایش‌های به $GL_2(\bar{\mathbb{F}}_p)$ تحویل‌ناپذیر باشند. نحوه‌ی استفاده از این قضیه را در بخش آخر توضیح می‌دهیم.

۷ دگردیسی‌های نمایش‌های گالوا

فرض کنید یک خم بیضوی روی اعداد گویا داریم و می‌خواهیم نمایش‌های گالوا روی آن را بررسی کنیم. در این صورت اگر نمایش روی کل مدول تیت را در نظر بگیریم، بررسی آن به نظر کار بسیار سختی است. پس ما ابتدا نمایش روی نقاط l–تابی را بررسی می‌کنیم و سعی می‌کنیم با ایده‌ای با بررسی این نمایش، کار را تمام کنیم!

در این قسمت نمایش‌های گالوا را کلی‌تر می‌نویسیم و سعی می‌کنیم شرط‌های محدودکننده‌ای روی آن قرار دهیم (شرط‌هایی که مطمئن باشیم برای نمایش‌های گالوا روی خم‌های نیمه‌پایدار درست باشند) و مدولاریتی آن‌ها را اثبات کنیم.

[41] Absolutely irreducible
[42] flat

تعریف ۲. *فرض کنید K یک توسیع متناهی $\mathbb{Q}_l$ باشد و O حلقه‌ی اعداد صحیح K باشد. رسته $\mathcal{C}_O$ را تعریف کنید: اشیا از همه Oـ جبرهای نوتری موضعی A (با ایدآل ماکسیمال) به همراه نگاشتی پوشا (نگاشت تشدید [43]) $\pi : A \to O$ تعریف کنید به‌طوری که $A/\mathfrak{m}_A \cong O/\mathfrak{m}_O = \mathbb{F}_q$ و نگاشت‌های بین این اشیا را برابر Oـ جبر هوموموفیسم‌های روی O تعریف کنید (نمودار زیر جابجا شود):*

$$\begin{array}{ccc} A & \longrightarrow & A' \\ {\scriptstyle \pi_A}\downarrow & & \downarrow{\scriptstyle \pi_{A'}} \\ O & \xrightarrow{\mathbb{1}_O} & O \end{array}$$

حال یک شیء از رسته بالا مثل A را فیکس کنید. می‌خواهیم دگردیسی [44] نمایش‌های گالوا را تعریف می‌کنیم:

تعریف ۳. ۱) *یک نمایش گالوا $\rho_0 : G_{\mathbb{Q}} \to GL_m(\mathbb{F}_q)$ فیکس کنید. منظور از یک بالابری [45] ρ از این نمایش گالوا یک نمایش $\rho : G_{\mathbb{Q}} \to GL_m(A)$ است به‌طوری که ρ به پیمانه‌ی $\mathfrak{m}_A$ برابر ρ_0 شود.*

۲) *دو بالابری ρ, ρ' از ρ_0 را اکیداً معادل [46] گوییم هرگاه یک ماتریس C وجود داشته باشد که به پیمانه‌ی $\mathfrak{m}_A$ همانی شود و داشته باشیم:*

$$\rho(g) = C^{-1}\rho'(g)C,\ g \in G_{\mathbb{Q}}$$

۳) *منظور از یک دگردیسی ρ_0 یک مولفه‌ی هم‌ارزی از رابطه‌ی هم‌ارزی بالا است.*

حال شرط‌هایی که گفتیم را باید روی دگردیسی‌های نمایش‌های گالوا قرار دهیم: پس یک نمایش گالوا $\rho_0 : G_{\mathbb{Q}} \to GL_m(\mathbb{F}_q)$ فیکس کنید و S را مجموعه‌ی اعداد اولی قرار دهید که ρ_0 در آن‌ها شاخه‌ای است. این یک مجموعه‌ی متناهی است (چرا؟). حال فرض کنید Σ یک مجموعه‌ی متناهی از اعداد اول باشد.

تعریف ۴. *دگردیسی ρ را از نوع D_Σ گوییم هرگاه:*

۱) *ρ بیرون از $\Sigma \cup S \cup \{p\}$ غیرشاخه‌ای باشد.*

۲) $\det \rho = \chi_p$.

۳) *برای هر $l \in S$،*

$$\rho|_{I_l} \sim \begin{bmatrix} 1 & * \\ 0 & 1 \end{bmatrix}$$

این شرط نیمه‌پایداری در l است.

۴) *تحدید $\rho|_{D_p}$ یا "صاف" باشد یا "معمولی" [47]. این‌ها دو شرط موضعی هستند تعریف آن‌ها از سطح این مقاله بالاتر است.*

همچنین به دگردیسی‌ای، مجاز [48] گوییم هرگاه برای Σیی D_Σ باشد. حال مجموعه‌های همه‌ی دگردیسی‌های از نوع D_Σ و دگردیسی‌های مدولار از نوع D_Σ را به ترتیب با $DA_\Sigma(A)$ و $DM_\Sigma(A)$ نشان می‌دهیم. در این صورت می‌توان ثابت کرد که این‌ها دو مجموعه‌ی متناهی هستند و اگر به آن‌ها به عنوان فانکتور [49] نگاه کنیم:

$$DM_\Sigma \subseteq DA_\Sigma : \mathcal{C}_O \to \mathbf{FiniteSets}$$

نمایش‌پذیر [50] خواهند بود. پس دو عضو از $\mathcal{C}_O$ وجود دارند مثل R_Σ و $\mathbb{T}_\Sigma$ به‌طوری که

$$DM_\Sigma(A) = \mathrm{Hom}(\mathbb{T}_\Sigma, A) \subseteq DA_\Sigma(A) = \mathrm{Hom}(R_\Sigma, A)$$

حال با قرار دادن $A = \mathbb{T}_\Sigma$ یک تابع کانونی $\phi_\Sigma : R_\Sigma \to \mathbb{T}_\Sigma$ پیدا می‌کنیم. حال می‌توانیم نمایش‌های زیر را پیدا می‌کنیم:

$$\rho_\Sigma^{univ} : G_{\mathbb{Q}} \to GL_2(R_\Sigma)$$

$$\rho_\Sigma^{univ.mod} : G_{\mathbb{Q}} \to GL_2(\mathbb{T}_\Sigma)$$

به‌طوری که با ϕ_Σ به هم مربوط می‌شوند. کاری که وایلز برای اثبات این قضیه می‌کند، اثبات قضیه زیر است:

قضیه ۵. (*قضیه اصلی*) *ϕ_Σ ایزومورفیسم در رسته $\mathcal{C}_O$ است.*

[43] Augmentation
[44] Deformation
[45] Lift
[46] Strictly equivalent
[47] Ordinary
[48] Admissible
[49] Functor
[50] Representable

این قضیه کار را تمام می‌کند، زیرا نشان می‌دهد که تعداد اعضای مجموعه‌های $DM_\Sigma(A)$ و $DA_\Sigma(A)$ برای همه‌ی Aها یکی است و پس همه‌ی دگردیسی‌های مجاز، مدولار هستند و پس قضیه مدولاریتی خم‌های نیمه‌پایدار به اتمام می‌رسد. نحوه‌ی اثبات وایلز به این شکل است که او روی تعداد اعضای Σ استقرا می‌زند و R_Σ و T_Σ را به‌طور صریح می‌سازد و انتخاب هوشمندانه‌ای برای توابع تشدید آن‌ها می‌کند و سپس با تکنیک‌های جبر این مسئله را به یک نامساوی درباره تعداد اعضای یک ناوردا در این حلقه‌ها تبدیل می‌کند و آن را اثبات می‌کند. شکافی که در اثبات ابتدایی وایلز وجود داشت این بود که اثبات پایه استقرا ($\Sigma = \emptyset$) کامل نبود و او و تیلور مجبور شدند یک سال دیگر روی آن زمان بگذارند تا آنرا اثبات کنند. این کار را باز با استفاده از تکنیک‌های جبری مشابه حالت گام استقرا اثبات کردند. برای خواندن اثبات دقیق آن می‌توانید به مقاله وایلز [۱] رجوع کنید.

۸ اتمام اثبات

پس فرض کردیم که معادله‌ی فرما یک جواب نابدیهی دارد: $A^p + B^p = C^p$ حال خم فری را به شکل زیر تعریف کردیم: $F : y^2 = x(x - A^p)(x + B^p)$ و نمایش‌های گالوای روی آن را در نظر گرفتیم. می‌توان دید که این نمایش‌ها مجاز هستند و پس از قضیه اصلی وایلز، مدولار هستند و پس یک فرم مدولار نرمال در $S_2(\Gamma_0(N))$ برای Nی پیدا می‌کنیم که N عددی خالی از مربع خواهد بود که $N \mid 2ABC$ (این از ترکیب قضیه‌ای از میزر [۵۱] و ریبت بدست می‌آید که آن را بیان نکردیم. می‌توانید این را در [۴] پیدا کنید). حال با استفاده از تفکیک‌کننده یک خم می‌توان شرطی روی نمایش‌های گالوای بدست آمده از نقاط lـ تابی قرار داد. اگر این کار را برای خم فری انجام دهیم نتیجه می‌شود که همه‌ی اعداد اولی که ABC را عاد می‌کنند در شروط قضیه‌ی ریبت صدق می‌کنند (به عنوان q در قضیه). پس طبق قضیه ریبت می‌توان فرم کاسپی نرمال در $S_2(\Gamma_0(2))$ پیدا کرد. اما داریم $S_2(\Gamma_0(2)) = \{0\}$ (این به‌سادگی از این‌که فشرده‌سازی $\mathbb{H}/\Gamma_0(2)$ گونه [۵۲] صفر دارد بدست می‌آید) و پس فرم مدولار نرمالی در آن وجود ندارد. این همان تناقضی بود که دنبالش بودیم!

مراجع

[1] Wiles A. Modular elliptic curves and Fermat's Last Theorem. 1995.

[2] Darmon H. Rational points on modular elliptic curves. 2003.

[3] Ribet K. From the Taniyama-Shimura conjecture to Fermat's last theorem. 1990.

[4] Panchishkin Alexei A. Manin, Yu. I. Introduction to Modern Number Theory. 2005.

[5] P. Samuel. Algebraic Theory of Numbers. 1970.

[6] Hungerford T. W. Algebra. Springer, 1974.

[51] B. Mazur
[52] Genus

مقدمه‌ای بر شبکه‌های عصبی(قسمت اول)

سید سروش هاشمی

چکیده

در این مقاله ابتدا مفاهیم اولیه یادگیری ماشین به ساده‌ترین شکل ممکن مطرح می‌شوند. سپس به بررسی شبکه‌های عصبی واقعی پرداخته می‌شود و تلاش می‌شود با الهام گرفتن از آن‌ها شبکه‌های عصبی مصنوعی طراحی کنیم. در نهایت نیز برخی انواع پیچیده شبکه‌های عصبی معرفی می‌شوند و تعدادی از کاربردهای شگفت‌انگیز شبکه‌های عصبی بیان می‌شوند.

۱ مقدمه

در طول تاریخ ریاضی‌دانان و دانشمندان بسیاری سعی کردند کارهایی که مختص انسان‌ها تلقی می‌شد را در ماشین‌ها شبیه سازی کنند یا به صورت قابل دفاعی به آن‌ها نسبت دهند. به طور مثال Alan Turing در سال ۱۹۳۷ ماشینی ساخت که توانایی استنتاج و تصمیم گیری بنابر استنتاج‌های خود را داشت(به تعبیری می‌توانست فکر کند و تصمیم بگیرد). همین طور John von Neumann در سال ۱۹۴۹ برنامه‌ای نوشت که توانایی بازنویسی کد خود از نو را داشت(به تعبیری می‌توانست مشابه ویروس‌ها تولید مثل کند). همچنین موفقیت‌های چشمگیری در افزایش درک ماشین‌ها از دنیای اطراف و ایجاد حواس پنجگانه در آن‌ها به دست آمده است. علاوه بر این رویکردهای عملی، ایده‌پردازان و فیلسوفان بسیاری درباره شباهت ماشین‌ها و انسان‌ها سخن گفته‌اند که یکی از جالب‌ترین آن‌ها ایده Samuel Butler درباره اعمال شدن نظریه داروین روی ماشین‌هاست. او معتقد بود همان طور که نظریه داروین برای موجودات زنده درست است، برای ماشین‌ها نیز صدق می‌کند و روزی می‌رسد که ماشین‌ها هویت(cognition) پیدا می‌کنند و کم کم جایگزین انسان‌ها می‌شوند.

در بین تمام این تلاش‌ها و ایده‌ها، یکی از پرچالش‌ترین و پرکاربردترین موضوعات، «یادگیری ماشین» بود. آیا یک ماشین می‌تواند چیزی بیاموزد؟ آیا می‌تواند همانند انسان تجربه کسب کند و اشتباهاتش را اصلاح کند؟ آیا یک ماشین می‌تواند ماشین‌های دیگر را آموزش دهد؟ آیا می‌توان برای یک ماموریت خطرناک(مانند نجات انسان‌ها از حریق، یا از کار انداختن یک بمب) تیمی از ماشین‌ها تربیت کرد که هر یک در زمینه‌ای متخصص باشند و بتوانند ماموریت را بدون به خطر افتادن جان انسان‌ها انجام دهند؟ «یادگیری» می‌تواند ماشین‌ها را به موجوداتی به مراتب کاربردی‌تر تبدیل کند.

۲ مفاهیم اولیه یادگیری ماشین

در مرحله اول باید به یک مدل ریاضی برای «یادگیری» دست یابیم. یکی از مدل‌های بسیار ساده و البته کاربردی به این صورت است:

«تابعی مانند $f : X \mapsto Y$ داریم که لزوما محاسبه پذیر نیست یا فرمولی برای محاسبه آن نداریم. مثلا یافتن نام یک

انسان با دیدن چهره او. تعدادی ورودی و خروجی نمونه مانند $(x^{(1)},y^{(1)}),(x^{(2)},y^{(2)}),\dots,(x^{(n)},y^{(n)})$ داریم. می‌خواهیم با کمک این نمونه‌ها تابعی مانند $\hat{f}:X\mapsto Y$ بیابیم که علاوه بر نمونه‌های موجود بتواند برای نمونه‌های جدید نیز خوب کار کند.»

علت این نوع از اندیس گذاری برای نمونه‌ها این است که خود x,y ممکن است ساختاری پیچیده‌تر از یک عدد داشته باشند. مثلا یک بردار یا ماتریسی از اعداد باشند. در این صورت نیاز داریم برای اشاره به درایه‌های ماتریس از اندیس‌های بیشتری استفاده کنیم. برای جلوگیری از ابهام، در این مقاله همواره اندیس‌های بالای متغیر نشانه اندیس نمونه است و اندیس پایین متغیر نشانه اندیس اجزاء (مثلا درایه‌ها) آن نمونه از آن متغیر. مثلا اگر x ساختار ماتریسی داشته باشد، $x^{(17)}_{3,5}$ نشان دهنده درایه سطر ۳ و ستون ۵ از نمونه شماره ۱۷ است.

احتمالا با دیدن این مدل تکنیک‌هایی مانند regression در آمار و علوم دیگر به ذهنتان خطور کرده است. با وجود سادگی، این تکنیک‌ها نیز جزو راه حل‌های این سوال محسوب می‌شوند. بنابراین شما می‌توانید از یک رگرسیون ساده به عنوان یک ابزار «یادگیری ماشین» استفاده کنید.

حال چگونه میزان شباهت یک تابع $\hat{f}$ با تابع f را بسنجیم؟

۱.۲ تابع loss

یک تعریف ساده و البته بسیار الهام بخش برای «میزان تفاوت تابع $\hat{f}$ و f » به این صورت است:

$$loss(f,\hat{f}):=\int_{x\in D_f} d(f(x),\hat{f}(x))$$

که در آن

$$d(f(x),\hat{f}(x))$$

تفاوت خروجی تابع f و $\hat{f}$ به ازای ورودی x است که بسته به نوع خروجی تابع f می‌توان آن را به صورت‌های مختلف تعریف کرد. به طور مثال اگر خروجی تابع f یک ماتریس ستونی باشد، می‌توان این تابع را به صورت

$$d(f(x),\hat{f}(x)):=(\hat{f}(x)-f(x))^T(\hat{f}(x)-f(x))$$

تعریف کرد.

حال اگر بخواهیم به ازای یک $\hat{f}$ داده شده، مقدار $loss(f,\hat{f})$ را محاسبه کنیم، چه کنیم؟ برای محاسبه این مقدار یک مشکل اساسی داریم. مشکل این است که تابع f را نداریم و نمی‌توانیم به ازای هر ورودی دلخواه $x\in D_f$ مقدار $f(x)$ را محاسبه کنیم و در فرمول تابع loss از آن استفاده کنیم. راه حل این مشکل، تخمین زدن تابع loss با مقدار آن در نقاط نمونه $(x_1,y_1),\dots,(x_n,y_n)$ است. بنابراین فرض می‌کنیم

$$loss(f,\hat{f})\approx\sum_{(x,y)\in\text{samples}} d(y,\hat{f}(x)) \qquad (۱)$$

تقریب خوبی برای تابع loss است. (البته برای این که این تقریب قابل قبول باشد باید نمونه‌ها و روش نمونه گیری چند خاصیت خوب، مثل عادلانه بودن نمونه گیری، را داشته باشند زیرا در غیاب این شروط بسیاری از نتایج آتی به خطر می‌افتند)

بنابراین سوال اولیه تبدیل به این سوال شد: «تابع $\hat{f}$ را بیابید که مقدار (تخمین) تابع loss را کمینه کند». این یک سوال بهینه سازی جالب و بسیار شایع در بسیاری از شاخه‌های علمی است.

۲.۲ مدل

چون در این سوال محدودیتی برای $\hat{f}$ قائل نشدیم، می‌توانیم به هر شکلی که بخواهیم این تابع را تعریف کنیم. بنابراین به جای این که به دنبال یک ضابطه برای این تابع بگردیم، این تابع را به صورت جفت‌های مرتب تعریف می‌کنیم به طوری که

$$(x^{(1)},y^{(1)}),\dots,(x^{(n)},y^{(n)})\in\hat{f}$$

باشد و مقدار $\hat{f}$ برای سایر مقادیر دامنه را به دلخواه مقداردهی می‌کنیم. این تابع مقدار تخمین loss را صفر می‌کند، بنابراین بهترین تابع است. اما در تعریف این تابع می‌توانیم به ورودی‌هایی که در نمونه‌ها نیستند هر مقدار خروجی دلخواه را نسبت دهیم، پس لزوما خروجی تابع $\hat{f}$ و f در سایر نقاط دامنه مشابه نیستند حتی اگر تخمین loss صفر باشد. این مشکل به این خاطر پیش آمد که ما به دنبال هیچ الگو و رابطه‌ای بین ورودی‌ها و خروجی‌های نمونه نگشتیم و فقط سعی کردیم تخمین تابع loss را کمینه کنیم. برای

حل این مشکل، تابع $\hat{f}$ را به دسته‌ای از توابع محدود می‌کنیم که حدس می‌زنیم می‌توانند الگو طبیعی بین ورودی و خروجی را بیابند و از آن برای تخمین مقدار f در سایر اعضای دامنه نیز استفاده کنند. این دسته از توابع را «مدل» می‌گویند. مثلا در رگرسیون خطی، فرض می‌کنیم تابع $\hat{f}$ به صورت

$$\hat{f}(\mathbf{x}) = \mathbf{w}\mathbf{x} + b = b + \sum_i w_i x_i$$

است پس مدل، تمام توابع خطی است. مدل می‌تواند خیلی ساده (مانند مثال قبل) یا بسیار پیچیده باشد. به عنوان یک مثال از یک مدل پیچیده می‌توان به مدلی که شامل توابعی به شکل زیر است اشاره کرد ($\mathbf{x}$ یک بردار ستونی با k درایه است) :

$$\begin{aligned}\hat{f}(\mathbf{x}) &= w_1x_1 + \cdots + w_kx_k \\ &+ w_{k+1}x_1x_2 + w_{k+2}x_1x_3 + \cdots + w_{\frac{k(k+1)}{2}}x_{n-1}x_n \\ &\vdots \\ &+ w_{2^k-1}x_1x_2\dots x_k\end{aligned} \qquad (۲)$$

این تابع در واقع تمامی 2^k ترکیب k درایه ورودی خود را به صورت وزن‌دار با هم جمع می‌کند.

باید مدل را طوری بسازیم که شامل خود تابع f یا یک تخمین خوب از آن باشد. برای این کار از دانش اولیه خود درباره مسئله الهام می‌گیریم. به طور مثال اگر باور داشته باشیم که یک ترکیب خطی از ورودی، خروجی را تولید می‌کند، کافی است مدل را به گونه‌ای تعریف کنیم که فقط توابع خطی را شامل شود. البته تلاش‌های جالبی برای یافتن خانواده‌ای از توابع که می‌توانند هر تابعی را تخمین بزنند انجام شده که یکی جالب‌ترین آن‌ها قضیه universal approximation [۲] است که ثابت می‌کند یک شبکه عصبی با تنها یک لایه مخفی [1] می‌تواند دسته بسیار بسیار بزرگی از توابع شایع در مسائل بسیاری از حوزه‌ها را تخمین بزند.

حال که تابع loss را تعریف کردیم و تابع $\hat{f}$ را به مدل محدود کردیم، باید بهترین تابع موجود در مدل را بیابیم که تخمین loss را کمینه کند. به الگوریتمی که بهترین تابع (یا یک تابع خوب) از بین توابع موجود در مدل را برای تخمین f می‌یابد، «بهینه‌ساز» [2] گویند.

۳.۲ الگوریتم‌های بهینه‌ساز

درباره این الگوریتم‌ها ذکر چند مسئله خالی از لطف نیست. هر الگوریتم بهینه‌ساز برای دسته خاصی از مدل‌ها طراحی شده است. بعضی اوقات مدل طوری تعریف شده که یافتن بهترین تابع برای تخمین f ممکن است. در این حالت الگوریتم‌های مختلف بر سر زمان یافتن آن با هم رقابت می‌کنند. اما در بسیاری مواقع، گستره توابع مدل به اندازه‌ای بزرگ و ناهموار می‌شود که الگوریتمی برای یافتن بهترین تابع $\hat{f}$ در آن وجود ندارد. در این شرایط علاوه بر زمان اجرا، کیفیت توابع پیدا شده نیز در مقایسه دو الگوریتم مورد نظر قرار می‌گیرند.

به طور خلاصه هر الگوریتم یادگیری ماشین از ۴ بخش تشکیل شده:

۱. داده‌های نمونه

۲. تعریف تابع loss

۳. مدل

۴. الگوریتم بهینه‌ساز

قبل از ورود به شبکه‌های عصبی، ذکر نکته‌ای دیگر درباره الگوریتم‌های یادگیری ماشین خالی از لطف نیست. فرض کنید ما ۱۰۰ عکس از چهره یک شخص در حالت‌های مختلف داریم و می‌خواهیم به یک ماشین آموزش دهیم که بتواند وجود یا عدم وجود چهره این شخص در عکس را تشخیص دهد. وقتی این ماشین آموزش داده شد، می‌خواهیم دقت آن را اندازه گیری کنیم. مثلا بگوییم این ماشین در ٪۹۰ مواقع درست کار می‌کند. برای اندازه گیری این دقت چه راه حلی دارید؟

[1] hidden
[2] optimizer

۴.۲ محاسبه دقت

یک راه ساده و البته الهام بخش برای اندازه گیری دقت، اینگونه است: «روی تمام اعضای دامنه، تابع $\hat{f}$ چند درصد جواب درست تولید می‌کند؟» اما برای محاسبه دقت با این تعریف، به خود تابع f نیاز داریم. مثلا در مثال عکس از چهره، باید بتوانیم تمام عکس‌های موجود را تولید کنیم و به ازای هر عکس تشخیص دهیم آیا چهره آن شخص در عکس وجود دارد یا نه و سپس نتیجه تشخیص خود را با خروجی تابع $\hat{f}$ مقایسه کنیم. این کار برای مسائلی که دامنه f در آن‌ها بزرگ است اصلا امکان پذیر نیست. حتی برای حالاتی که دامنه f خیلی بزرگ نیست، اصلا صرفه زمانی و اقتصادی ندارد. برای حل این مشکل مجددا سعی می‌کنیم این مقدار را تخمین بزنیم. برای تخمین زدن دقت، کافی است چند نمونه‌ی دیگر از دامنه به همراه خروجی تابع f در آن نمونه‌ها را داشته باشیم و فقط خروجی f و $\hat{f}$ را در آن نمونه‌ها مقایسه کرده و دقت را تخمین بزنیم. اگر برای محاسبه دقت از همان نمونه‌هایی که با آن‌ها $\hat{f}$ را یافته‌ایم استفاده کنیم قطعا تعداد پاسخ‌های غلط بسیار کمی خواهیم داشت (زیرا $\hat{f}$ طوری طراحی شده که تخمین تابع loss را کمینه کند) و در نتیجه دقت محاسبه شده بالا خواهد بود. اما اگر برای محاسبه دقت از داده‌های جدیدی استفاده کنیم، تخمین ما از دقت، مقدار واقعی‌تری را نشان می‌دهد. به همین خاطر قبل از شروع فاز یادگیری، معمولا داده‌های موجود را به نسبت ٪۸۰ به ٪۲۰ به دو دسته «داده‌های یادگیری» [3] و «داده‌های تست» [4] تقسیم می‌کنند و در فاز یادگیری فقط از داده‌های یادگیری و در فاز تست (محاسبه دقت) فقط از داده‌های تست استفاده می‌کنند. البته هیچ نکته خاصی در اعداد ۸۰ و ۲۰ وجود ندارد و شما می‌توانید به هر نسبتی که خودتان می‌پسندید داده‌هایتان را تقسیم کنید. البته در مواقعی که داده‌های کمی موجود است از ایده‌های دیگری استفاده می‌کنند.

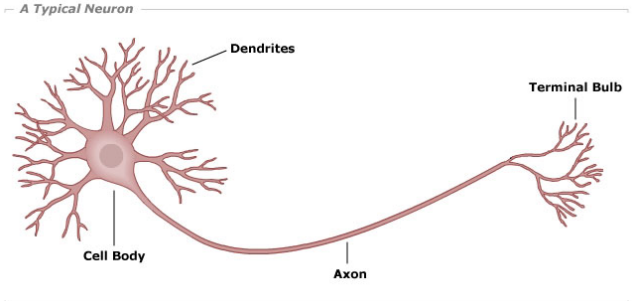
حال که جنبه‌های مختلف یک الگوریتم یادگیری ماشین را دیدیم به بررسی شبکه‌های عصبی می‌پردازیم.

۳ شبکه‌های عصبی

ابتدا کمی شبکه‌های عصبی بدن موجودات زنده را بررسی می‌کنیم. سپس تلاش می‌کنیم با الهام گرفتن از آن‌ها مدل ریاضی تولید کنیم که یک تابع را محاسبه می‌کند. نهایتا این مدل را به شبکه‌های عصبی واقعی پیچیده‌تر نزدیک می‌کنیم و مدل ریاضی تولید شده را ارتقا می‌دهیم.

۱.۳ شبکه‌های عصبی در علوم زیستی

شبکه عصبی در بدن یک موجود زنده از تعداد زیادی نورون تشکیل شده است که هر یک تعدادی سیگنال ورودی دریافت کرده و یک سیگنال خروجی تولید می‌کند. به تعبیری یک نورون تابعی از سیگنال‌های ورودی خود را به عنوان خروجی تولید می‌کند. خروجی هر نورون می‌تواند به تعداد دلخواهی نورون منتقل شود. (ورودی و خروجی نورون‌ها یک پالس الکتریکی است. در محاسبه خروجی نورون، تنها ولتاژ این پالس‌های الکتریکی تاثیرگذار است. می‌توان نورون را به صورت یک تابع از ولتاژ پالس‌های الکتریکی ورودی به ولتاژ پالس الکتریکی خروجی دید. پس اگر خروجی یک نورون را به صورت موازی به چند نورون ارسال کنیم ولتاژ ثابت می‌ماند و فقط جریان تغییر می‌کند) یک شبکه عصبی متشکل از تعداد زیادی نورون است که ورودی هر یک، خروجی تعداد دلخواهی نورون دیگر است. بنابراین یک شبکه عصبی معادل یک گراف جهت‌دار بزرگ از نورون‌هاست. ذکر این نکته جالب است که شبکه‌های عصبی با دور نیز مشاهده شده‌اند.

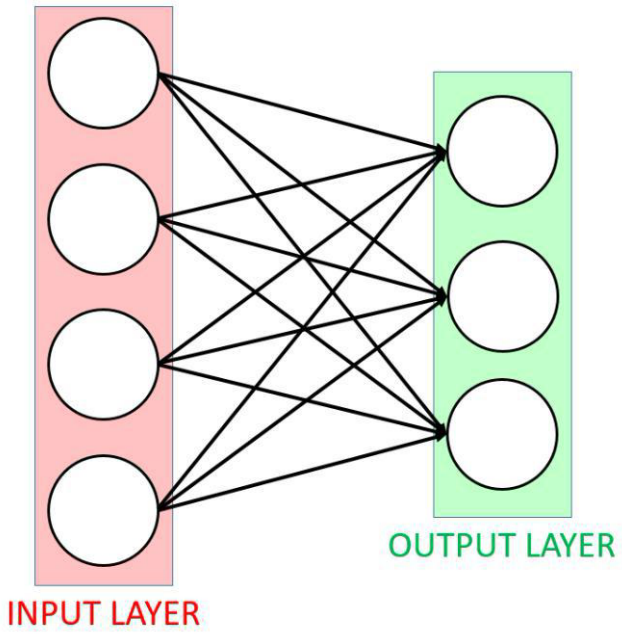


شکل ۱.۳

[3] train set
[4] test set

۲.۳ شبکه‌های عصبی بدون لایه مخفی

فرض کنید تابع f یک ماتریس ستونی با ۴ درایه را می‌گیرد و یک ماتریس ستونی با ۳ درایه تولید می‌کند. برای شبیه‌سازی این تابع با نورون‌ها می‌توانیم به ازای هر درایه خروجی تابع یک نورون در نظر بگیریم که تمام ۴ درایه ورودی به آن وصل هستند. هر نورون باید یک تابع از ۴ ورودی خود را حساب کند و خروجی دهد. برای ساده شدن مدل ریاضی فرض می‌کنیم هر نورون یک ترکیب خطی از ورودی‌های خود را محاسبه می‌کند. در این صورت مدل تبدیل به شکل زیر می‌شود. در این شکل برای هر یال یک «وزن» تعریف می‌کنیم که در واقع همان ضریب درایه متناظر آن در ورودی برای محاسبه درایه مورد نظر خروجی است. مثلا یالی که درایه اول ورودی را به درایه دوم خروجی وصل می‌کند وزن درایه اول در ترکیب خطی متناظر با درایه دوم خروجی است.



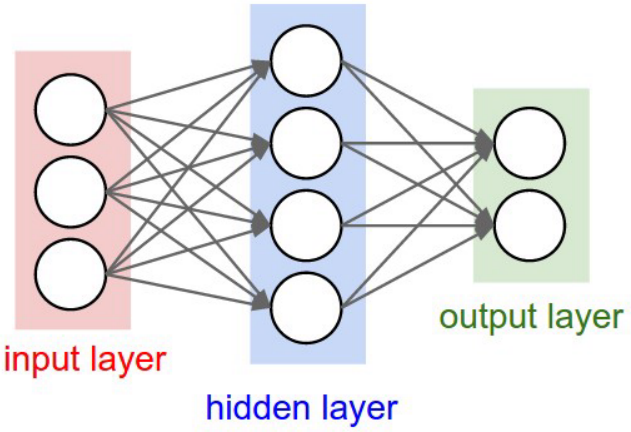
شکل ۲.۳

این مدل ریاضی به صورت جالب و ساده‌ای تبدیل به یک ضرب ماتریسی می‌شود. اگر ماتریس ورودی را $\mathbf{x}$ و ماتریس خروجی را $\mathbf{y}$ بگیریم می‌توان تعریف کرد:

$$\mathbf{y} = \mathbf{W}\mathbf{x}$$

که $\mathbf{W}$ یک ماتریس با ۳ سطر و ۴ ستون است که درایه‌های سطر i ام آن ضرایب ترکیب خطی متناظر با درایه i ام خروجی است.

در شبکه‌های عصبی واقعی، نورون‌ها ورودی خود را از تعدادی نورون می‌گیرند و خروجی خود را به تعدادی نورون دیگر می‌دهند. همچنین تابعی که نورون‌ها محاسبه می‌کنند که ترکیب خطی ساده نیست. در ادامه تلاش می‌کنیم مدل خود را کمی پیچیده‌تر کنیم تا این ساختار را شبیه سازی کنیم. برای این کار ابتدا فرض می‌کنیم یک لایه از نورون‌ها داریم که با محاسبه ترکیب خطی ورودی‌ها لایه‌ای میانی از اعداد تولید می‌کنند. سپس لایه‌ای دیگر از نورون‌ها، خروجی این لایه میانی را گرفته و با محاسبه تابعی از این اعداد، اعداد درایه‌های خروجی را تولید می‌کنند.



شکل ۲.۳

اگر تابعی که نورون‌های لایه دوم محاسبه می‌کنند نیز یک ترکیب خطی از ورودی‌های خود باشد، درایه‌های ماتریس خروجی کل این شبکه عصبی، ترکیب خطی ورودی‌های لایه اول آن می‌شود. پس اگر قصد پیچیده‌تر شدن مدل را داریم باید تابع دیگری برای این لایه از نورون‌ها انتخاب کنیم. فرض می‌کنیم نورون‌های موجود در لایه دوم این شبکه عصبی، همگی تابع غیر خطی و مشابه g را محاسبه می‌کنند. بنابراین مدل جبری این شبکه عصبی به صورت زیر است (این مدل برای شکل ۲.۳ است. در این مثال، $\mathbf{W}$ یک ماتریس با ۴ سطر و ۳ ستون است) :

$$\hat{f}(\mathbf{x}) = \begin{bmatrix} g(\mathbf{W}\mathbf{x}) \\ g(\mathbf{W}\mathbf{x}) \end{bmatrix} \qquad (۳)$$

ممکن است سوال کنید چون تابع g و ورودی $\mathbf{x}$ برای هر ۲ درایه ماتریس خروجی یکی است، آیا این ۳ مساوی نمی‌شوند؟ این بستگی به تعریف تابع g دارد. ممکن است تابع g طوری تعریف شود که وابسته به اندیس درایه خروجی خود باشد. مثلا نورونی که درایه دوم ماتریس خروجی را تولید می‌کند تابعی از درایه سوم و چهارم لایه میانی باشد و نورونی که درایه اول ماتریس خروجی را تولید می‌کند تابعی از دو درایه اول لایه میانی باشد. همچنین ممکن است تابع g پارامتر دیگری علاوه بر x داشته باشد که این پارامتر برای دو نورون ماتریس خروجی متفاوت باشد. همچنین روش‌هاش دیگری برای تمایز دادن بین توابع این دو درایه وجود دارد.

می‌توان ایده اضافه کردن لایه میانی را بیش از یکبار استفاده کرد و شبکه‌های عصبی با لایه‌های میانی بیشتر ساخت.

۳.۳ شبکه‌های عصبی عمیق

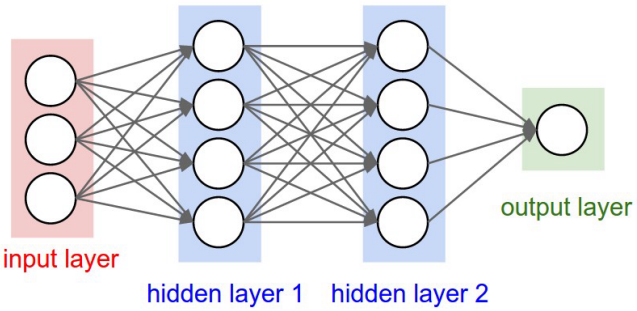
یک شبکه عصبی عمیق از تعدادی لایه میانی، تشکیل شده که خروجی هر لایه تابعی از خروجی لایه قبلی و پارامترهای آن لایه است. مثلا در مثال قبل، لایه اول تابعی از ورودی شبکه عصبی و ماتریس $\mathbf{W}$ بود. همچنین لایه دوم تابعی از خروجی لایه اول و پارامترهای تابع g . یک شبکه عصبی عمیق به صورت

$$\hat{f}(x) = f_n(f_{n-1}(\dots f_2(f_1(x,\theta_1),\theta_2)\dots),\theta_{n-1}),\theta_n)$$

تعریف می‌شود که برای وضوح بیشتر می‌توان آن را به صورت زیر نوشت:

$$\begin{aligned} y_1 &= f_1(x,\theta_1) \\ y_2 &= f_2(y_1,\theta_2) \\ &\vdots \\ y_{n-1} &= f_{n-1}(y_{n-2},\theta_{n-1}) \\ y_n &= f_n(y_{n-1},\theta_n) \end{aligned} \qquad (۴)$$

لایه اول (تابع f_1) ورودی x را گرفته و تابعی بر حسب آن و پارامترهای خود (θ_1) را محاسبه می‌کند. لایه دوم، خروجی لایه اول (y_1) را گرفته و تابعی بر حسب آن و پارامترهای خود (θ_2) محاسبه می‌کند. همین طور تا لایه آخر که تابعی بر حسب خروجی لایه قبلی (y_{n-1}) و پارامترهای خود (θ_n) را محاسبه می‌کند و به عنوان خروجی کل شبکه عصبی می‌دهد.



شکل ۳.۳

به تعداد توابع این شبکه عصبی، عمق آن می‌گویند. برد هر تابع میانی به دلخواه طراح شبکه عصبی است ولی برد تابع آخر باید مشابه برد تابع f باشد. به تعداد درایه‌های ماتریس خروجی هر تابع میانی، عرض آن لایه می‌گویند. تحقیقات جالبی درباره ارتباط عمق و عرض شبکه‌های عصبی با پیچیدگی توابع تولید شده توسط آن‌ها انجام شده است که یکی از جذاب‌ترین آن‌ها [۳] است. برای طرح یکی از نتایج جالب این مقاله لازم است یک اصطلاح پر کاربرد در تئوری شبکه‌های عصبی را معرفی کنم.

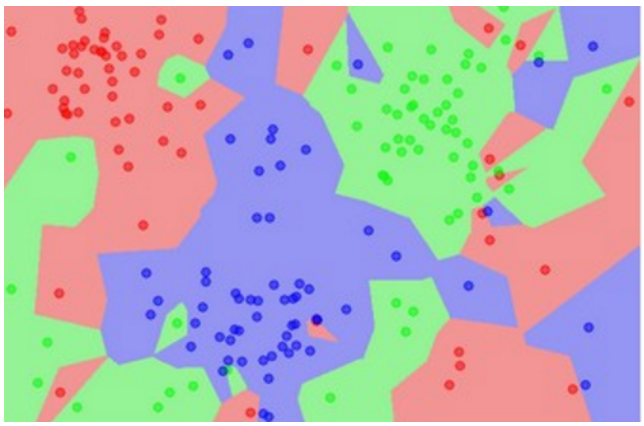
۱.۳.۳ ظرفیت یک شبکه عصبی

برای یک مدل، «ظرفیت [۵] » مدل، به طور شهودی، گستره توابع موجود در مدل و پیچیدگی آن‌ها تعریف می‌شود. هر چه تعداد و پیچیدگی توابع موجود در یک مدل بیشتر باشد، ظرفیت مدل بیشتر است. مثلا مدلی که فقط شامل توابع خطی است نمی‌تواند یک تابع درجه دو را نمایش دهد بنابراین ظرفیت آن کم است. اما مدلی که علاوه بر توابع خطی، توابع درجه دو را نیز شامل می‌شود دارای ظرفیت بیشتری است. برای محاسبه کمّی ظرفیت مدل، روش‌های زیادی وجود دارد که یکی از آن‌ها «بعد VC [۶] » است. این معیار، ظرفیت یک مدل را این گونه تعریف می‌کند:

[5] capacity
[6] VC dimension

«فرض می‌کنیم برد تابع f فقط مجموعه $\{0,1\}$ باشد. یک مجموعه دلخواه $x_1, x_2, \ldots, x_n$ در نظر می‌گیریم. اگر بتوانیم به ازای هر 2^n حالت $y_1, y_2, \ldots, y_n$ یک تابع مانند g در مدل بیابیم که $\forall i \in \{1, \ldots, n\} : \ g(x_i) = y_i$ می‌گوییم این مدل می‌تواند مجموعه $x_1, x_2, \ldots, x_n$ را به صورت دلخواه پخش کند. حال اگر مدل توانایی پخش کردن دلخواه هر مجموعه n تایی از ورودی‌ها را داشته باشد، می‌گوییم مدل ظرفیت پخش دلخواه n ورودی دلخواه را دارد. بزرگترین عدد n که مدل توانایی پخش دلخواه n ورودی دلخواه را داشته باشد، ظرفیت مدل در بعد VC است.»

می‌توان این تعریف را به حالات پیچیده‌تری نیز تعمیم داد. مثلا فرض کنید سبز,آبی,قرمز $\in y$ و $\mathbf{x}$ یک نقطه در صفحه $\mathbf{R}^2$ باشد و نمونه‌ها در شکل ۱.۳.۳ با نقطه مشخص شده باشند. در این صورت تابعی که در شکل ۱.۳.۳ ناحیه‌های صفحه را رنگ آمیزی کرده، نمونه‌ها را به درستی در مجموعه‌های «قرمز»، «آبی» و «سبز» پخش کرده است.



شکل ۱.۳.۳

در مقاله [۳] ، فرمولی برای تعداد ناحیه‌هایی که یک خانواده شایع از شبکه‌های عصبی می‌تواند آن‌ها را از هم تمیز دهد (مشابه شکل ۱.۳.۳) ارائه شده است. در این فرمول فرض شده عرض تمام لایه‌های میانی یکسان و برابر w است. همچنین عمق شبکه عصبی (تعداد لایه‌های میانی آن) برابر d و تعداد درایه‌های ماتریس ورودی کل شبکه عصبی n در نظر گرفته شده. تابع غیر خطی لایه‌های میانی نیز تابعی در نظر گرفته شده که مجموعه شایعی از شبکه‌های عصبی را شامل شود. در این مقاله ثابت شده تعداد ناحیه‌های خطی قابل تمیز این خانواده از شبکه‌های عصبی، از فرمول زیر به دست می‌آید.

$$O\left(\binom{w}{n}^{n(d-1)} - w^n\right) \qquad (۵)$$

این فرمول نشان می‌دهد اضافه کردن d بسیار بیشتر از اضافه کردن سایر مشخصه‌های شبکه عصبی می‌تواند ظرفیت شبکه عصبی را افزایش دهد. به همین خاطر شبکه‌های عصبی عمیق بیشتر از شبکه‌های عصبی عریض مورد توجه و استفاده قرار گرفته‌اند.

۴.۳ شبکه‌های عصبی پیچیده‌تر

یکی از مشاهدات جالب دانشمندان علوم شناختی، شبکه‌های عصبی‌ای بود که دور داشتند، یعنی خروجی یک نورون به نورون دیگری فرستاده می‌شد و آن نورون خروجی‌اش را به نورون دیگری می‌فرستاد و در نهایت این سلسله مجددا به نورون اولیه برمی‌گشت. محققان شبکه‌های عصبی مصنوعی تلاش‌هایی برای شبیه‌سازی این نوع شبکه‌های عصبی واقعی کردند [۷] . همچنین شبکه‌های عصبی مصنوعی طراحی شد که دارای حافظه بودند (مانند ماشین تورینگ اتفاقاتی که تا به این لحظه برای آن‌ها رخ داده است را به شکلی نگهداری می‌کردند) [۸] . همچنین شبکه‌های عصبی متنوعی برای کاربردهای خاص با معماری‌ها و توابع میانی خاص طراحی شدند و نتایج شگفت‌آوری تولید کردند. به طور مثال شبکه‌های عصبی پیچشی [۹] که برای تحلیل عکس‌ها مورد استفاده قرار می‌گیرند. یادگیری عمیق در حال حاضر یکی از فعال‌ترین و بزرگترین جامعه محققان را دارد و هر سال تعداد بسیار زیادی مقاله منتشر می‌شوند که هر کدام مسیر جدید برای استفاده از شبکه‌های عصبی در مسائل واقعی گشوده‌اند.

[7] Recursive Neural Networks
[8] Recurrent Neural Networks, Neural Turing Machines
[9] Convolutional Neural Networks

۴ کاربردهای شگفت‌انگیز شبکه‌های عصبی

همان طور که در این مقاله خواندید شبکه‌های عصبی از کنار هم قرار گرفتن تعداد زیادی ایده ساده تشکیل شده‌اند و هر زمان که سادگی یکی از این ایده‌ها، دقت پاسخ یا سرعت اجرای کد را به اندازه غیر قابل چشم‌پوشی‌ای تحت تاثیر قرار می‌داد، ایده‌هایی به عنوان جایگزین مطرح می‌شدند. این روند همچنان ادامه دارد و هر ساله، تغییرات شگرفی در ایده‌های اولیه شبکه‌های عصبی ایجاد می‌شود و همچنین ایده‌های جدید نیز دستخوش تغییراتی می‌شوند. جامعه بزرگ و توانمندی از دانشمندان و مهندسان حوزه علوم کامپیوتر و علوم دیگر از جمله علوم زیستی، هر سال با کمک شبکه‌های عصبی به پیشرفت‌های شگفت‌آوری در حوزه‌های مختلف دست می‌یابند. تشخیص بیماری‌هایی از جمله سرطان، تشخیص تخلف رانندگی، تشخیص چهره(با دقتی فراتر از انسان)، تبدیل گفتار به نوشتار و نوشتار به گفتار، نقاشی کردن، خلق یک اثر موسیقی، امنیت، تشخیص الگوهای بازار بورس، رنگی کردن عکس‌های سیاه و سفید، استخراج صدا از فیلم‌های صامت فقط با استفاده از لرزش حاصل از موج‌های مکانیکی، تشخیص سن و جنسیت از روی عکس، تولید اتوماتیک caption برای عکس‌ها و ... تعداد کمی از کاربردهای شگفت‌انگیز شبکه‌های عصبی هستند.

یکی از استفاده‌های بسیار شگفت‌انگیز شبکه‌های عصبی توانایی آن‌ها در یادگیری بازی‌های ساده کامپیوتری از جمله قارچ‌خور است. می‌توان شبکه عصبی طراحی کرد که بتواند بدون راهنمایی و فقط با بازی کردن بازی قارچ‌خور یاد بگیرد چطور بازی کند که امتیاز بیشتری کسب کند (دقیقا مشابه انسان). در واقع شما بازی را اجرا می‌کنید و شبکه عصبی شروع به بازی کردن می‌کند. در ابتدا حتی نمی‌داند باید به سمت راست حرکت کند که برنده شود. اما بعد از مدتی و با آزمون و خطا متوجه می‌شود (مانند انسان). بعد از آن شروع به مشاهده و یادگیری اجزا مختلف بازی می‌کند. مثلا ابتدا مستقیما به دل یک دشمن می‌زند و می‌بازد اما بعد از مدتی متوجه می‌شود که این کار موجب باخت می‌شود. همین طور در اواسط بازی خود را درون چاه‌ها می‌اندازد ولی به مرور متوجه می‌شود چطور از روی چاه‌ها بپرد. جالب این جاست که بعد از حدود ۳۰۰ دقیقه بازی کردن، سطح بازی این شبکه عصبی از انسان نیز فراتر می‌رود و حرکات ترکیبی شگرفی انجام می‌دهد که از توان محاسباتی انسان خارج است.

۵ ادامه دارد ...

در قسمت بعدی این مقاله تعدادی از الگوریتم‌های بهینه‌ساز شبکه‌های عصبی را بررسی می‌کنیم و برخی از توابع استفاده شده در شبکه‌های عصبی را معرفی و تجزیه و تحلیل خواهیم کرد.

مراجع

[1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. deep learning. deeplearningbook.org.

[2] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. Neural networks, 2(5):359–366, 1989.

[3] Guido F Montufar, Razvan Pascanu, Kyunghyun Cho, and Yoshua Bengio. On the number of linear regions of deep neural networks. In Advances in neural information processing systems, pages 2924–2932, 2014.

تبدیل فوریه در گروه‌ها

امیرحسین اخلاصی

۱ مقدمه

تبدیل فوریه[1] یکی از معروفترین و پرکاربردترین ابزارهای ریاضی است که توسط بسیاری از ریاضی‌دانان، فیزیک‌دانان و مهندسان به کار می‌رود. تحلیل فوریه در واقع یک ابزار خوب برای تحلیل خواص یک تابع از جمله امواج و سیگنال‌ها است. برای مثال در تحلیل سیگنال‌ها، تبدیل فوریه‌ی یک تابع نشان‌دهنده‌ی فرکانس‌های موجود در آن تابع و در کوانتوم، تبدیل فوریه‌ی یک تابع موج مکان، تابع موج تکانه را به ما می‌دهد. در حل معادلات دیفرانسیل نیز به علت خواص خوب تبدیل فوریه، از آن استفاده می‌شود. در احتمال و ... می‌توان حضور تبدیل فوریه را دریافت.

اما در باب چیستی تبدیل فوریه، باید گفت که نشان‌دهنده‌ی آن است که یک تابع چه جور ترکیب خطی‌ای از توابع به شکل $e^{2\pi ix}$ است. یعنی یک موج از ترکیب خطی امواج با فرکانس‌های مختلف به وجود آمده است و تبدیل فوریه می‌خواهد به ما نشان دهد که ضرایب این ترکیب خطی چگونه هستند. برای مثال اگر $f(x)=\sum_{n\in\mathbb{Z}} c_n e^{2\pi inx}$ یک تابع با تناوب ۱ باشد و $\hat{f}(n)=\int_0^1 e^{-2\pi inx}f(x)dx$ سری فوریه f باشد، آنگاه $\hat{f}(n)=c_n$. البته این گزاره تحت شرایطی ممکن است برقرار نباشد و شرایط لازم به قضیه‌ی وارون فوریه[2] معروف است.

اما ما در مسائلمان معمولاً با چهار نوع تبدیل فوریه مواجه می‌شویم:

۱. **تبدیل فوریه:** $f:\mathbb{R}\to\mathbb{C}$

(۱)
$$\hat{f}(\xi)=\int_{-\infty}^{+\infty} e^{-2\pi ix\xi}f(x)dx,\quad f(x)=\int_{-\infty}^{+\infty} e^{2\pi ix\xi}\hat{f}(\xi)d\xi$$

۲. **سری فوریه:** $f:\mathbb{R}\to\mathbb{C}$ با تناوب ν

(۲)
$$\hat{f}(n)=\frac{1}{\nu}\int_0^{\nu} e^{-\frac{2\pi ixn}{\nu}}f(x)dx,\quad f(x)=\sum_{n=-\infty}^{\infty}\hat{f}(n)e^{\frac{2\pi ixn}{\nu}}$$

۳. **تبدیل فوریه‌ی گسسته‌زمان:** $f:\mathbb{Z}\to\mathbb{C}$

(۳)
$$\hat{f}(x)=\sum_{n=-\infty}^{\infty} f(n)e^{-2\pi inx},\quad f(n)=\int_0^1 e^{2\pi inx}\hat{f}(x)dx$$

۴. **تبدیل فوریه‌ی گسسته:** $f:\mathbb{Z}\to\mathbb{C}$ با تناوب N

(۴)
$$\hat{f}(m)=\frac{1}{N}\sum_{n=0}^{N-1} f(n)e^{\frac{-2\pi imn}{N}},\quad f(n)=\sum_{m=0}^{N-1}\hat{f}(m)e^{\frac{2\pi imn}{N}}$$

یکی از سؤالاتی که می‌خواهیم به آن جواب دهیم این است که آیا می‌توان ادبیات مشترکی برای هر چهار نوع تبدیل پیدا کرد؟ درواقع این سؤال وقتی بیشتر مطرح می‌شود که بدانیم خواص تبدیل فوریه در تمام چهار مورد فوق شبیه هم است. از طرف دیگر وقتی پای

[1] Fourier transform

[2] Fourier inversion theorem

تابع دلتا به میان می‌آید، مسائل بیشتری برای ریاضی‌دانان مطرح می‌شود. زیرا تابع دلتا از لحاظ ریاضی بی‌معنی است و صرفاً یک نماد فیزیکی است که فقط زیر انتگرال معنا می‌دهد. به عنوان مثال در تبدیل فوریه‌ی پیوسته به پیوسته، برای تابع $f(x) = e^{2\pi ixax}$ تبدیل فوریه‌ی آن موجود نمی‌باشد، یعنی $\int_{-\infty}^{+\infty} e^{-2\pi ix\xi} f(x)dx$ موجود نیست. این در حالی‌ست که

$$\int_{-\infty}^{+\infty} e^{2\pi ix\xi}\delta(\xi - a)d\xi = e^{2\pi ixax} = f(x) \qquad (۵)$$

اما همان طور که می‌دانیم، تابع دلتا در واقع وجود خارجی ندارد و صرفاً مفهومی است که توسط فیزیک‌دانان توسعه یافته است. در واقع تابع دلتا نمادی برای نمایش اندازه‌ی $d\mu(x) = \delta(x-a)dx$ است که در آن

$$\mu(E) = \begin{cases} 0 & a \in E \text{ اگر} \\ 1 & a \notin E \text{ اگر} \end{cases} \qquad (۶)$$

پس تبدیل فوریه چیزی فرای یک تبدیل از توابع به توابع است. در این مقاله سعی شده است چیستی تبدیل فوریه مورد بحث قرار گیرد. در سه بخش اول ابتدا به بیان مقدمات و پیش‌نیازها می‌پردازیم و بخش‌های بعدی را به بیان تبدیل فوریه و مسائل مورد نیاز اختصاص می‌دهیم.

۲ قضیه‌ی استون ـ وایرشتراس

فرض کنید X یک فضای توپولوژیک موضعاً فشرده و هاوسدرف باشد. فرض کنید $C(X,\mathbb{C})$ مجموعه‌ی تمام توابع پیوسته‌ی $f : X \to \mathbb{C}$ باشد.[۳] $C(X,\mathbb{C})$ را می‌توان به توپولوژی فشرده‌ـباز مجهز کرد که در آن مفهوم همگرایی $f_n \to f$ به معنای آن است که f_n روی همه‌ی زیرمجموعه‌های فشرده‌ی X به طور یکنواخت به f همگرا باشد. همچنین فرض کنید $C_c(X,\mathbb{C})$ مجموعه‌ی تمام توابع $f \in C(X,\mathbb{C})$ با محمل فشرده باشد.

تعریف ۱. برای تابع $f : X \to \mathbb{C}$ می‌گوییم f در بی‌نهایت صفر می‌شود هرگاه برای هر $\epsilon > 0$ ، مجموعه‌ی فشرده‌ی K وجود داشته باشد که برای xهای خارج از K داشته باشیم $|f(X)| < \epsilon$.

مجموعه‌ی تمام توابع پیوسته‌ی مختلط‌مقدار که در بی‌نهایت صفر می‌شوند را با $C_0(X,\mathbb{C})$ نشان می‌دهیم. واضح است که $C_c(X,\mathbb{C}) \subset C_0(X,\mathbb{C}) \subset C(X,\mathbb{C})$ ؛ در واقع $C_0(X,\mathbb{C}) = \overline{C_c(X,\mathbb{C})}$. همچنین می‌توان روی $C_0(X,\mathbb{C})$ نرم L^∞ را به شکل $\|f\|_\infty = \sup_{x\in X}|f(x)|$ تعریف کرد. مجموعه‌ی $A \subset C(X,\mathbb{C})$ را یک جبر خودالحاق می‌گویند هرگاه تحت جمع، ضرب، ضرب اسکالر و مزدوج‌گیری بسته باشد. می‌گوییم جبر A هیچ جا صفر نمی‌شود هرگاه همه‌ی اعضای A همزمان در یک نقطه صفر نشوند. می‌گوییم جبر A نقاط X را جدا می‌کند هر گاه برای هر دو نقطه‌ی متمایز $x, y \in X$ تابع $f \in A$ وجود داشته باشد که $f(x) \neq f(y)$.

قضیه ۱ (استون‌ـوایرشتراس[۴]). *اگر $A \subset C_0(X,\mathbb{C})$ یک جبر خودالحاق باشد که هیچ جا صفر نشود و نقاط X را جدا کند، در این صورت A در $C_0(X,\mathbb{C})$ چگال است.*

برای اطلاعات بیشتر در این زمینه به [۲] مراجعه شود.

۳ نظریه‌ی اندازه

در این قسمت به ارائه‌ی خلاصه‌ای از موارد مورد نیاز در نظریه‌ی اندازه می‌پردازیم.

۱.۳ اندازه

اندازه اولین بار توسط لبگ مطرح شد که به هر زیرمجموعه از $\mathbb{R}$ عددی را نسبت داد که به آن اندازه‌ی مجموعه می‌گفت به طوری که خاصیت جمع‌پذیری را دارا باشد و تحت انتقال ناوردا باشد. وی دو مفهوم اندازه‌ی داخلی و خارجی را معرفی کرد و مجموعه‌هایی را که برای آن‌ها این دو عدد مساوی می‌شدند مجموعه‌های اندازه‌پذیر نامید. در ابتدا معلوم نبود که آیا مجموعه‌ی اندازه‌ناپذیر وجود دارد

[۳] تمام تعاریف و قضایای مربوط به این بخش با جایگزین کردن $\mathbb{R}$ به جای $\mathbb{C}$ نیز برقرار است.

[4] Stone-Weierstrass theorem

یا نه ولی چندی بعد ویتالی مثالی از چنین مجموعه‌ای ساخت. کار لبگ برای تعریف اندازه‌ی لبگ روی $\mathbb{R}$ و $\mathbb{R}^n$، بعدها تعمیم پیدا کرد و اندازه‌های انتزاعی روی هر فضایی قابل تعریف شد. بخصوص اندازه‌ی احتمال روی فضاهای احتمال بسیار مورد توجه قرار گرفت.

اندازه را روی گردایه‌ای از زیرمجموعه‌های یک فضا که به σـجبر معروف است تعریف می‌کنند. یک σـجبر مانند Σ از مجموعه‌ی X گردایه‌ای از زیرمجموعه‌های X است که شامل X و $\emptyset$ است و تحت متمم‌گیری و اجتماع شمارا بسته باشد. به اعضای Σ اصطلاحاً مجموعه‌های اندازه‌پذیر می‌گویند.

تعریف ۲. یک اندازه روی (X,Σ) تابعی مانند $\mu : \Sigma \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ می‌باشد که دارای خواص زیر است:

۱. $\mu(\emptyset) = 0$

۲. **(σـجمع‌پذیری[5])** اگر $A_1, A_2, \ldots$ حداکثر شمارا مجموعه‌ی دو به دو مجزا از Σ باشند، داشته باشیم

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n) \qquad (۷)$$

اگر X یک فضای توپولوژیک باشد، به σـجبر تولیدشده با مجموعه‌های باز آن، σـجبر بُرل می‌گوییم و به اندازه‌های روی آن، اندازه‌ی بُرل می‌گوییم. در نظریه‌ی اندازه مجموعه‌های اندازه‌صفر (مجموعه‌هایی که $\mu(E) = 0$) قابل اغماض هستند و می‌توان در گزاره‌ها به آن‌ها توجه نکرد. برای مثال اگر دو تابع از X همه جا غیر از یک مجموعه‌ی اندازه‌صفر با هم برابر باشند، ما آن دو تابع را در نظریه‌ی اندازه یکی در نظر می‌گیریم و می‌گوییم این دو تابع تقریباً همه جا با هم برابرند. اندازه‌ی μ را کامل می‌گوییم هرگاه برای هر مجموعه‌ی اندازه‌صفر A، اگر $B \subset A$ ، آنگاه B نیز اندازه‌پذیر (و در نتیجه اندازه‌صفر) باشد. اگر μ یک اندازه روی (X,Σ) باشد، می‌توان با اضافه کردن این گونه مجموعه‌ها به Σ، آن را کامل کرد. (واضح است که برای همه‌ی اندازه‌ها نمی‌توان همزمان این کار را انجام داد.) مجموعه‌های اندازه‌پذیر روی $\mathbb{R}^n$ کامل‌شده‌ی σـجبر برل آن به توسط اندازه‌ی لُبگ[6] هستند. اندازه‌ی لبگ تحت انتقال و دوران ناوردا است. برای $\mathbb{R} \cup \{-\infty, \infty\}$ نیز می‌توان مجموعه‌های اندازه‌پذیر و اندازه‌ی لبگ را به راحتی تعمیم داد.

اما مهمترین خاصیتی که اندازه‌ها برای بسیاری از کارها مانند تعریف انتگرال و قضایای گوناگون باید دارا باشند، σـ متناهی بودن است؛ یعنی حداکثر شمارا مجموعه‌ی $A_1, A_2, \ldots$ موجود باشند که

$$X = \bigcup_n A_n \quad \text{و} \quad \forall i,\ \mu(A_i) < \infty \qquad (۸)$$

برای جزئیات بیشتر به [۴] مراجعه شود.

در قسمت آخر به تعریف جلو دادن[7] یک اندازه با یک تابع و توابع ناوردا اشاره می‌کنیم.

تعریف ۳. فرض کنید (X,Σ_1) و (Y,Σ_2) دو فضای اندازه باشند. در این صورت تابع $f : X \to Y$ را اندازه‌پذیر می‌نامیم هرگاه برای هر $E \in \Sigma_2$، $f^{-1}(E) \in \Sigma_1$. اگر μ یک اندازه روی (X,Σ_1) باشد، آنگاه می‌توان اندازه‌ی $f_*\mu$ را روی (Y,Σ_2) به صورت $f_*\mu(E) = \mu(f^{-1}(E))$ تعریف کرد. تابع اندازه‌پذیر $f : X \to X$ را μـ ناوردا می‌گوییم هرگاه $f_*\mu = \mu$.

۲.۳ انتگرال لبگ

از مهمترین ارکان نظریه‌ی اندازه انتگرال لبگ است؛ برای تابع اندازه‌پذیر $f : X \to \mathbb{R} \cup \{-\infty, \infty\}$ می‌توان

$$\int_X f d\mu \qquad (۹)$$

را تعریف کرد که تعمیمی از جمع وزن‌دار مقادیر تابع f روی X با وزن μ است. البته اگر تابع f هر دوی مقادیر منفی و مثبت را اتخاذ کند، ممکن است انتگرال فوق تعریف‌ناپذیر باشد. لذا ابتدا برای توابع نامنفی f، انتگرال تعریف می‌شود که مقدار آن می‌تواند ∞ باشد. در قدم بعدی تابع f را که می‌تواند مقادیر منفی را هم اتخاذ کند انتگرال‌پذیر می‌گوییم هرگاه $\int_X |f| d\mu < \infty$. اگر f

[5] σ- additivity

[6] Lebesgue measure

[7] push forward

انتگرال‌پذیر باشد، انتگرال f را به شکل

$$\int_X f d\mu = \int_X f^+ d\mu - \int_X f^- d\mu \qquad (۱۰)$$

تعریف می‌کنند که در آن $f^+ = \max(0, f)$ و $f^- = -\min(0, f)$.

انتگرال لبگ تمام خواص خوبی که از آن انتظار می‌رود، مانند خطی بودن و نامساوی مثلث را دارا است. مهمتر از آن اگر 1_A تابع مشخصه‌ی مجموعه‌ی اندازه‌پذیر A باشد، آنگاه

$$\int_X 1_A d\mu = \mu(A) \qquad (۱۱)$$

برای جزئیات بیشتر از خواص انتگرال لبگ به [۴] مراجعه کنید.

تذکر. اگر $E \subset X$ زیرمجموعه‌ی اندازه‌پذیری از X باشد، آنگاه

$$\int_E f d\mu = \int_X 1_E f d\mu \qquad (۱۲)$$

تا اینجا انتگرال را برای توابع اندازه‌پذیر حقیقی‌مقدار تعریف کردیم. اما به راحتی می‌توان این تعریف را برای توابع مختلط‌مقدار تعمیم داد. اگر $f = u + iv$ یک تابع اندازه‌پذیر مختلط‌مقدار باشد، f را انتگرال‌پذیر می‌گوییم هرگاه $\int_X |f| d\mu < \infty$. در صورتی که f انتگرال‌پذیر باشد، انتگرال آن را می‌توان به شکل

$$\int_X f d\mu = \int_X u d\mu + i \int_X v d\mu \qquad (۱۳)$$

تعریف کرد. انتگرال برای توابع مختلط نیز دارای تمام خواص خوب اشاره‌شده در بالا می‌باشد. نکته‌ی دیگری که می‌توان به آن اشاره کرد خاصیت خوب انتگرال در جلو دادن است. اگر $L : X \to Y$ تابعی اندازه‌پذیر باشد، آنگاه رابطه‌ی

$$\int_Y f d(L_* \mu) = \int_X (f \circ L) d\mu \qquad (۱۴)$$

برای توابع μـانتگرال‌پذیر f برقرار است. تا اینجا با خواص خوب انتگرال لبگ آشنا شدیم، اما در همه‌ی شرایط انتگرال لبگ رفتار مورد انتظار را ندارد و همان طور که در بخش قبل اشاره شد برای بعضی از خواص مورد انتظار نیاز است μ، σـمتناهی باشد.

قضیه ۲. *فرض کنید μ یک اندازه‌ی σـ متناهی باشد و $f \geq 0$ یک تابع اندازه‌پذیر باشد. در این صورت*

(آ) *اگر $\int_X f d\mu < \infty$ آنگاه $f < \infty$ a.e.*.

(ب) *اگر $\int_X f d\mu = 0$ آنگاه $f = 0$ a.e.*.

حال به تعریف فضای L^p می‌پردازیم.

تعریف ۴. برای $1 \leq p < \infty$ نرم L^p برای توابع اندازه‌پذیر به شکل

$$\|f\|_p = \left(\int_X |f|^p d\mu \right)^{\frac{1}{p}} \qquad (۱۵)$$

تعریف می‌شود. تابع f را L^p می‌گوییم هرگاه $\|f\|_p < \infty$. $L^p(X, \mu)$ مجموعه‌ی تمام توابع L^p است که در آن دو تابع را که تقریباً همه جا با هم برابر باشند را یکی در نظر می‌گیریم.

می‌توان نشان داد که$L^p(X, \mu)$ یک فضای برداری کامل (فضای باناخ) با نرم $\| \ \|_p$ می‌باشد. معمولاً مجموعه‌ی $L^p(X, \mu)$ را بر حسب استفاده به اختصار با $L^p(X)$ یا $L^p(\mu)$ نمایش می‌دهند. همگرایی L^p نیز به این معنا است که $\|f_n - f\|_p \to 0$.

قضیه ۳ (نامساوی هولدر[۸]). *فرض کنید $1 \leq p, q \leq \infty$ اعدادی باشند به طوری که $\frac{1}{p} + \frac{1}{q} = 1$، در این صورت برای هر دو تابع $f \in L^p(X, \mu)$ و $g \in L^q(X, \mu)$،*

$$\|f\|_p \|g\|_q \geq \|fg\|_1 \qquad (۱۶)$$

در آخر به قضایای همگرایی انتگرال اشاره می‌کنیم. می‌گوییم دنباله‌ی f_n تقریباً همه جا به f همگرا است و می‌نویسیم $f_n \to f$ a.e. هرگاه برای تمام مقادیر x به غیر از حداکثر یک مجموعه‌ی اندازه‌صفر $f_n(x) \to f(x)$. سؤال این جاست که در چه صورت $\int f_n d\mu \to \int f d\mu$؟ دو قضیه‌ی زیر به این موضوع پرداخته است.

قضیه ۴ (قضیه‌ی همگرایی یکنوا). *اگر f_n توابع نامنفی اندازه‌پذیر باشند به طوری که $f_n \to f$ a.e. و $f_n(x) \leq f(x)$ a.e. در این صورت*

$$\lim_{n \to \infty} \int_X f_n d\mu = \int_X f d\mu \qquad (۱۷)$$

[8] Holder's inequality

قضیه ۵ (قضیه‌ی همگرایی تسلطی). *فرض کنید f_n دنباله‌ای از توابع اندازه‌پذیر باشند که تحت تسلط تابع $g \in L^p(X,\mu)$ باشند. یعنی $|f_n| \leq |g|$ a.e. (در نتیجه f_nها نیز L^p هستند.) در این صورت اگر $f_n \to f$ a.e. آنگاه*

(آ)

$$\lim_{n\to\infty} \|f_n - f\|_p = 0 \qquad (۱۸)$$

یعنی $f_n \xrightarrow{L^p} f$.

(ب) *در حالت $p=1$ از (۱۸) نتیجه می‌شود که*

$$\int_X f_n d\mu \to \int_X f d\mu \qquad (۱۹)$$

۳.۳ قضیه‌ی فوبینی ـ تونلی

زمانی که با انتگرال‌های دوگانه یا چندگانه کار می‌کنیم، در ابتدا از یک مؤلفه و در ادامه از مؤلفه‌ی دیگر انتگرال می‌گیریم. همین کار را می‌توان برای اندازه‌ها کرد. اگر (X_1,Σ_1) و (X_2,Σ_2) دو فضای اندازه باشند، می‌توان σـجبر $\Sigma_1 \times \Sigma_2$ را که از مجموعه‌هایی به شکل $A \times B$ که $A \in \Sigma_1$ و $B \in \Sigma_2$ تولید می‌شود در نظر گرفت. حال اگر μ_1 و μ_2 دو اندازه‌ی مختلط به ترتیب روی (X_1,Σ_1) و (X_2,Σ_2) باشند، می‌توان اندازه‌ی $\mu_1 \times \mu_2$ را روی $(X_1 \times X_2, \Sigma_1 \times \Sigma_2)$ تعریف کرد به طوری که

$$\mu_1 \times \mu_2(A \times B) = \mu_1(A)\mu_2(B) \qquad (۲۰)$$

البته برای یکتایی نیاز است که μ_1 و μ_2 هر دو σـمتناهی باشند. (در نتیجه $\mu_1 \times \mu_2$ نیز σـمتناهی خواهد بود.)

قضیه‌ی فوبینی ـ تونلی[9] در واقع ترکیبی از دو قضیه‌ی فوبینی و تونلی است که برای انتگرال گرفتن دوگانه و چندگانه از توابع استفاده می‌شود.

قضیه ۶ (تونلی). *فرض کنید μ_1 و μ_2 دو اندازه‌ی σـمتناهی به ترتیب روی (X_1,Σ_1) و (X_2,Σ_2) باشند و f یک تابع اندازه‌پذیر نامنفی روی $X_1 \times X_2$ باشد. فرض کنید*

$$g(x) = \int_{X_2} f(x,y) d\mu_2(y) \qquad (۲۱)$$

در این صورت g یک تابع اندازه‌پذیر روی X_1 است و

(۲۲)

$$\int_{X_1} g(x) d\mu_1(x) = \int_{X_1\times X_2} f(x,y) d(\mu_1 \times \mu_2)(x,y)$$

قضیه ۷ (فوبینی). *فرض کنید μ_1 و μ_2 دو اندازه‌ی σـمتناهی به ترتیب روی (X_1,Σ_1) و (X_2,Σ_2) باشند و f یک تابع $\mu_1 \times \mu_2$ـ انتگرال‌پذیر روی $X_1 \times X_2$ باشد. در این صورت توابع $f^x(y) = f(x,y)$ روی X_2 تقریباً برای همه‌ی xها μ_2ـ انتگرال‌پذیر است و اگر*

$$g(x) = \int_{X_2} f(x,y) d\mu_2(y) \ \ a.e. \qquad (۲۳)$$

آنگاه g یک تابع اندازه‌پذیر و μ_1ـ انتگرال‌پذیر روی X_1 است و

(۲۴)

$$\int_{X_1} g(x) d\mu_1(x) = \int_{X_1\times X_2} f(x,y) d(\mu_1 \times \mu_2)(x,y)$$

در واقع وقتی می‌خواهیم انتگرال دوگانه بگیریم ابتدا لازم است که بدانیم آیا تابع انتگرال‌پذیر است که این را با استفاده از قضیه‌ی تونلی (قضیه ۶) بررسی می‌کنیم، سپس قضیه‌ی فوبینی (قضیه ۷) به ما کمک می‌کند تا انتگرال دوگانه بگیریم. البته در این مقاله خیلی به جزئیات توجه نمی‌کنیم و محک انتگرال‌پذیری توابع را به عهده‌ی خواننده می‌گذاریم.

۴.۳ قضیه‌ی رادون ـ نیکدیم

در حساب دیفرانسیل، با عباراتی مانند $dy = f(x)dx$ روبه‌رو میشویم. معنای این جمله معمولاً این است که برای هر تابع h، $\int h(y)dy = \int h(x)f(x)dx$. حال این سؤال پیش می‌آید که آیا این صورت‌بندی را می‌توان برای اندازه‌ها ادامه داد؟ اگر μ یک اندازه روی (X,Σ) و f یک تابع اندازه‌پذیر نامنفی باشد، آنگاه اندازه‌ی ν را می‌توان به شکل

$$\nu(E) = \int_E f d\mu \qquad (۲۵)$$

[9] Fubini-Tonelli theorem

تعریف کرد. در این حالت می‌توان نشان داد که برای هر تابع اندازه‌پذیر h،

$$\int_X h(x)d\nu(x) = \int_X h(x)f(x)d\mu(x) \qquad (۲۶)$$

لذا می‌نویسیم $d\nu = fd\mu$. حال این سؤال پیش می‌آید که برای دو اندازه‌ی μ و ν، چه موقع می‌توان رابطه‌ای به شکل $d\nu = fd\mu$ نوشت. واضح است که در این حالت اگر E یک مجموعه‌ی μـ اندازه‌صفر باشد، آنگاه νـ اندازه‌صفر نیز هست.

تعریف ۵. فرض کنید μ و ν دو اندازه روی (X,Σ) باشند. در این صورت می‌گوییم ν نسبت به μ پیوسته‌ی مطلق است و می‌نویسیم $\nu \ll \mu$، هرگاه هر مجموعه‌ی μـ اندازه‌صفر، νـ اندازه‌صفر نیز باشد.

واضح است که اگر $d\nu = fd\mu$، در این صورت $\nu \ll \mu$ اما قضیه‌ی زیر عکس این گزاره را بیان می‌کند:

قضیه ۸ (قضیه‌ی رادون‌ـ نیکدیم[۱۰]). *فرض کنید μ و ν دو اندازه‌ی σـ متناهی روی (X,Σ) باشند به طوری که $\nu \ll \mu$. در این صورت تابع اندازه‌پذیر $f : X \to [0,\infty)$ موجود است که $d\nu = fd\mu$؛ یعنی*

$$\nu(E) = \int_E f d\mu \qquad \forall E \in \Sigma \qquad (۲۷)$$

و این f تقریباً یکتا است؛ به این معنا که اگر g تابع دیگری با خواص فوق باشد، آنگاه $f = g$ a.e..

۵.۳ اندازه‌ی علامت‌دار و اندازه‌ی مختلط

در بسیاری از مواقع، وقتی ترکیب خطی مقادیر یک تابع را حساب می‌کنیم، ضرایب می‌توانند منفی یا حتی مختلط باشند، لذا لازم است که مفهوم اندازه را به اندازه‌های علامت‌دار یا حتی مختلط‌مقدار تعمیم داد.

تعریف ۶. فرض کنید (X,Σ) یک فضا همراه با یک σـ جبر باشد. در این صورت

(آ) $\mu : \Sigma \to \mathbb{R} \cup \{-\infty,\infty\}$ را یک اندازه‌ی علامت‌دار می‌نامیم، هرگاه $\mu(\emptyset) = 0$ و μ دارای خاصیت σـ جمع‌پذیری باشد. (رجوع شود به تعریف ۲)

(ب) $\mu : \Sigma \to \mathbb{C}$ را یک اندازه‌ی مختلط می‌نامیم، هرگاه $\mu(\emptyset) = 0$ و μ دارای خاصیت σـ جمع‌پذیری باشد.

تذکر. در قسمت σـ جمع‌پذیری اندازه‌های علامت‌دار و مختلط، معادله‌ی (۷) به ما می‌گوید جمع سمت راست، همگرایی مطلق است. لذا اندازه‌های علامت‌دار، تنها قادرند حداکثر یکی از مقادیر ∞ و $-\infty$ را اختیار کنند.

تعریف علامت‌دار و مختلط‌مقدار از اندازه‌ها به ما کمک می‌کند تا اندازه‌ها را با هم جمع و تفریق کنیم و فضای اندازه‌ها تبدیل به یک فضای برداری شود. همچنین انتگرال‌گیری در این فضای برداری خاصیت خطی داشته باشد. یعنی اعضا در رابطه‌ی

$$\int_X fd(\mu_1 + c\mu_2) = \int_X fd\mu_1 + c\int_X fd\mu_2 \qquad (۲۸)$$

صدق کنند. البته تعریف انتگرال برای اندازه‌های علامت‌دار و مختلط دارای ریزه‌کاری‌هایی است که در این جا به قسمتی از آن اشاره می‌کنیم.

تعریف ۷. فرض کنید μ یک اندازه‌ی علامت‌دار یا مختلط روی (X,Σ) باشد و برای هر $E \in \Sigma$، $S(E)$ مجموعه‌ی تمام دنباله‌های حداکثر شمارا از اعضای Σ مانند $(E_n)_n$ باشد که E_n ها دو به دو مجزا باشند و $\bigcup_n E_n = E$. در این صورت اندازه‌ی تغییرات μ یا قدر مطلق μ را به شکل

$$|\mu|(E) = \sup\left\{\sum_n |E_n| \mid (E_n)_n \in S(E)\right\} \qquad (۲۹)$$

تعریف می‌کنیم و تغییرات کل μ را برابر با $|\mu|(X) = \|\mu\|$ می‌گیریم.

می‌توان نشان داد که $|\mu|$ یک اندازه است که برای هر $E \in \Sigma$، $|\mu(E)| \le |\mu|(E)$. همچنین اگر μ یک اندازه‌ی مختلط باشد $\|\mu\| < \infty$. $|\mu|$ به ما کمک می‌کند تا انتگرال‌پذیری یک تابع را

[10] Radon-Nikodym theorem

تعریف کنیم. تابع اندازه‌پذیر f را μـانتگرال‌پذیر می‌گوییم هرگاه $|\mu|$ـانتگرال‌پذیر باشد.

تذکر. اگر μ یک اندازه علامت‌دار یا مختلط باشد و A یک مجموعه‌ی اندازه‌پذیر باشد که $\mu(A) = 0$ ، این به معنای اندازه‌صفر بودن A (ناچیز بودن A) نیست. زیرا ممکن است A از دو قسمت تشکیل شده باشد که اندازه‌ی یک قسمت مثبت و اندازه‌ی قسمت دیگر منفی باشد و جمع آنها صفر شود. لذا A را μـاندازه‌صفر می‌نامیم هرگاه برای هر مجموعه‌ی اندازه‌پذیر $B \subseteq A$، $\mu(B) = 0$ یا به طور معادل $|\mu|(A) = 0$.

اما برای تعریف انتگرال روی اندازه‌های حقیقی‌مقدار، نیاز به قضیه‌ی زیر است.

قضیه ۹ (تجزیه‌ی هانـ ژردان[11]). *فرض کنید μ یک اندازه‌ی علامت‌دار روی (X, Σ) باشد. در این صورت اندازه‌های یکتای μ^+ و μ^- وجود دارند که $\mu = \mu^+ - \mu^-$ و μ^+ و μ^- نسبت به هم تکینه هستند؛ یعنی مجموعه‌های اندازه‌پذیر P و N موجود باشند که $X = P \cup N$ و $P \cap N = \emptyset$ و P یک مجموعه‌ی μ^-ـاندازه‌صفر و N یک مجموعه‌ی μ^+ـاندازه‌صفر باشد.*

در قضیه‌ی قبل واضح است که $|\mu| = \mu^+ + \mu^-$. حال با استفاده از قضیه ۹ می‌توان انتگرال را برای اندازه‌های علامت‌دار تعریف کرد. درواقع ایده‌ی اصلی تعریف استفاده از رابطه‌ی مورد انتظار (۲۸) برای اندازه‌ها است. اگر f یک تابع μـانتگرال‌پذیر باشد، انتگرال f را به شکل

$$\int_X f d\mu = \int_X f d\mu^+ - \int_X f d\mu^- \qquad (۳۰)$$

تعریف می‌کنیم. همچنین اگر μ یک اندازه‌ی مختلط باشد، μ را می‌توان به طور یکتا به شکل $\mu = \mu_1 + i\mu_2$ نوشت که μ_1 و μ_2 اندازه‌های علامت‌دار هستند. در این صورت برای تابع μـانتگرال‌پذیر f، انتگرال f را می‌توان به شکل

$$\int_X f d\mu = \int_X f d\mu_1 + i \int_X f d\mu_2 \qquad (۳۱)$$

تعریف کرد. انتگرال برای اندازه‌های علامت‌دار و مختلط دارای خواص خوب مورد انتظار مانند خطی بودن نسبت به توابع و صحت رابطه‌ی (۲۸) است. برای اطلاعات بیشتر از خواص انتگرال به [۷] مراجعه کنید.

اندازه‌های علامت‌دار و مختلط در بسیاری از قضایای مهمی که در قبل ذکر شدند، صدق می‌کنند؛ در قضیه‌ی همگرایی تسلطی (قضیه ۵) و قضیه‌ی فوبینی (قضیه ۷) می‌توان اندازه‌ها را علامت‌دار یا مختلط در نظر گرفت. همچنین در صورت قضیه‌ی رادونـ نیکدیم (قضیه ۸) نیز می‌توان ν را مختلط در نظر گرفت. از طرفی می‌توان نشان داد برای دو اندازه‌ی مختلط μ و ν که در رابطه‌ی $\nu = f\mu$ صدق می‌کنند، رابطه‌ی

$$|d\nu| = |f||d\mu| \qquad (۳۲)$$

برقرار است. علاوه بر این ها برای توابع مختلط می‌دانیم $\mu \ll |\mu|$، در نتیجه با استفاده از قضیه‌ی رادونـ نیکدیم و رابطه‌ی (۳۲) می‌توان نتیجه گرفت که تابع $\theta : X \to \mathbb{R}$ موجود است که $d\mu = e^{i\theta}|d\mu|$ که از این معادله می‌توان به عنوان راه حل جایگزینی برای تعریف انتگرال برای اندازه‌های مختلط استفاده کرد.

۶.۳ اندازه‌ی منظم

فرض کنید X یک فضای توپولوژیک هاوسدرف و موضعاً فشرده باشد و $\mathscr{B}$ σـجبر برل آن باشد.

تعریف ۸. اندازه‌ی برل μ را روی $(X, \mathscr{B})$ منظم می‌گوییم هرگاه

۱. برای هر مجموعه‌ی فشرده K، $\mu(K) < \infty$،

۲. برای هر مجموعه‌ی برل E،

$$\mu(E) = \inf\{\mu(U) \mid E \subseteq U, \text{ باز } U\} \qquad (۳۳)$$

۳. برای هر مجموعه‌ی برل E که $\mu(E) < \infty$،

$$\mu(E) = \sup\{\mu(K) \mid K \subseteq E, \text{ فشرده } K\} \qquad (۳۴)$$

اگر μ یک اندازه‌ی علامت‌دار یا مختلط باشد، آن را منظم می‌نامیم هرگاه $|\mu|$ منظم باشد.

[11] Hahn-Jordan decomposition theorem

در اینجا بررسی فضای $L^p(X,\mu)$ برای اندازه‌های برل، خصوصاً اندازه‌های منظم خالی از لطف نیست. توپولوژی روی فضای L^p با نرم $\| \ \|_p$ تعریف می‌شود. (تعریف ۴) همچنین برای $p=\infty$، $L^\infty(X)$ فضای توابع کران‌دار با نرم

$$\|f\|_\infty = \sup\{f(x) \mid x \in X\} \qquad (۳۵)$$

است که توپولوژی آن همان توپولوژی همگرایی یکنواخت است. اگر $\mu\| < \infty$ (تعریف ۷) آنگاه $L^q \subseteq L^p$ برای $1 \leq p \leq q \leq \infty$ و توپولوژی L^q ظریفتر از توپولوژی L^p خواهد بود. در نتیجه

$$C_0(X,\mathbb{C}) \subseteq L^\infty(X) \subseteq L^p_\mu(X). \qquad (۳۶)$$

اگر μ یک اندازه‌ی منظم باشد، در این صورت $C_c(X,\mathbb{C})$ یک زیرمجموعه‌ی چگال L^p برای هر $1 \leq p < \infty$ است.

از طرفی دیگر توابع مشخصه‌ی مجموعه‌های برل را می‌توان با توابع پیوسته به صورت نقطه به نقطه تقریب زد. برای حالت خاص $p=1$ می‌توان از توضیحات فوق استفاده کرد و قضیه‌ی زیر را نتیجه گرفت.

قضیه ۱۰. *فرض کنید X یک فضای توپولوژیک موضعاً فشرده و هاوسدرف باشد.*

(آ) *فرض کنید مجموعه‌ی $A \subseteq C_0(X,\mathbb{C})$ با نرم $\| \ \|_\infty$ در $C_0(X,\mathbb{C})$ چگال باشد. همچنین فرض کنید μ_1 و μ_2 دو اندازه‌ی برل مختلط روی X باشند به طوری که*

$$\int_X f d\mu_1 = \int_X f d\mu_2 \qquad \forall f \in A \qquad (۳۷)$$

در این صورت $\mu_1 = \mu_2$.

(ب) *فرض کنید μ_1 و μ_2 دو اندازه‌ی منظم روی X باشند به طوری که*

$$\int_X f d\mu_1 = \int_X f d\mu_2 \qquad \forall f \in C_c(X,\mathbb{C}) \qquad (۳۸)$$

در این صورت $\mu_1 = \mu_2$.

تابع $\psi : \mathbb{C}(X,\mathbb{C}) \to \mathbb{C}$ را مثبت می‌گوییم هرگاه برای هر تابع $f \geq 0$، $\psi(f) \geq 0$.

قضیه ۱۱ (قضیه‌ی نمایش ریس‌ـ مارکف‌ـ کاکوتانی[۱۲]). *فرض کنید $T : C_c(X,\mathbb{C}) \to \mathbb{C}$ یک تابع خطی باشد. در این صورت*

(آ) *اگر T مثبت باشد، آنگاه اندازه‌ی منظم یکتای μ وجود دارد که*

$$T(f) = \int_X f d\mu\,, \qquad \forall f \in C_c(X,\mathbb{C}) \qquad (۳۹)$$

(ب) *اگر $\|T\| < \infty$، آنگاه اندازه‌ی مختلط منظم یکتای μ وجود دارد که*

$$T(f) = \int_X f d\mu\,, \qquad \forall f \in C_c(X,\mathbb{C}) \qquad (۴۰)$$

۴ گروه‌های آبلی موضعاً فشرده

یک گروه توپولوژیک گروهی است مانند G با یک ساختار توپولوژیک هاوسدرف که در آن تابع ضرب $(x,y) \to xy$ و تابع وارون $x \to x^{-1}$ پیوسته باشند. در این نوشته مجال پرداختن به گروه‌های توپولوژیک نیست و صرفاً به خواصی از گروه‌های توپولوژیک آبلی موضعاً فشرده پرداخته می‌شود. برای اطلاع بیشتر از خواص گروه‌های توپولوژیک به [۳] مراجعه شود.

۱.۴ اندازه‌ی هار

یکی از ویژگی‌هایی که اندازه‌ی لبگ روی $\mathbb{R}^n$ دارد این است که تحت انتقال ناوردا است. حال این سؤال پیش می‌آید که آیا می‌توان مشابه این حرف را به طور کلی برای گروه‌ها زد؟ جواب برای گروه‌های موضعاً فشرده مثبت است. برای هر $y \in G$ فرض کنید $L_y(x) = yx$ و $R_y(x) = xy$ به ترتیب انتقال از چپ و انتقال از راست تحت y باشند و $\mathrm{inv}\,(x) = x^{-1}$ تابع وارون باشد. با توجه به تعریف ۳ می‌خواهیم اندازه‌ی برل ناصفر μ روی G را طوری پیدا کنیم که تحت تمامی این عملگرها ناوردا باشد.

[12] Riesz-Markov-Kakutani representation theorem

قضیه ۱۲ (هار[۱۳]). *اگر G گروه موضعاً فشرده باشد، آنگاه اندازه‌ی منظم (تعریف ۸) غیر صفر μ موجود است که ناوردا تحت انتقال راست باشد. یعنی برای هر $y \in G$، $(R_y)_* \mu = \mu$. در ضمن این μ با تقریب ضریب ثابت ناوردا است. یعنی اگر ν اندازه‌ی دیگری باشد که خواص فوق را دارا باشد، آنگاه $\nu = c\mu$ که c ضریبی ثابت است. همچنین $\mu(G) < \infty$ اگر و فقط اگر G فشرده باشد.*

به اندازه‌ی تعریف‌شده در قضیه‌ی بالا اندازه‌ی هار[۱۴] می‌گویند. در قضیه‌ی بالا اندازه‌ی هار برای انتقال راست تعریف شد، که متأسفانه لزوماً تحت انتقال چپ ناوردا نیست. اما اگر G آبلی باشد، انتقال چپ و راست یک معنا را می‌دهند و همچنین این اندازه تحت وارون گرفتن نیز ناوردا می‌ماند.

یکی از راه‌هایی که می‌توان قضیه‌ی هار را ثابت کرد این است که تابع خطی مثبتی مانند $T : C_c(G, \mathbb{C}) \to \mathbb{C}$ بسازیم که تحت انتقال راست ناوردا باشد. در این صورت طبق قضیه‌ی نمایش ریس‌ـ مارکف‌ـ کاکوتانی (قضیه‌ی ۱۱) می‌توان یک اندازه‌ی هار به T نسبت داد. در قضیه‌ی ۲۲ از این روش استفاده شده است.

اگر G σـ همبند باشد (یعنی حداکثر شمارا مؤلفه‌ی همبندی داشته باشد) اندازه‌ی هار آن σـ متناهی خواهد بود. برای اطلاعات بیشتر از خواص اندازه‌ی هار به [۱] مراجعه کنید.

تعریف ۹. اگر X یک فضای متریک باشد، در این صورت تابع $f : G \to X$ را پیوسته‌ی یکنواخت می‌گوییم هرگاه برای هر $\epsilon > 0$ همسایگی V از عضو خنثای G موجود باشد که $d(f(x), f(y)) < \epsilon$ هرگاه $xy^{-1} \in V$.

قضیه ۱۳. *فرض کنید $f \in L^p(G)$. اگر $f_x(y) = f(yx^{-1})$، در این صورت تابع*

$$x \to f_x \qquad (G \to L^p(G)) \qquad (۴۱)$$

پیوسته‌ی یکنواخت است.

۲.۴ گروه دوگان

در این بخش فرض کنید G یک گروه آبلی موضعاً فشرده باشد و عمل آن را با $+$ نشان می‌دهیم. همچنین گروه دایره‌ی واحد S^1 را به عنوان زیرگروهی از $\mathbb{C}\backslash\{0\}$ با عمل ضرب در نظر بگیرید.

تعریف ۱۰. یک مشخصه‌ی G یک همریختی $\xi : G \to S^1$ است. اگر $\widehat{G}$ را برابر با همه‌ی مشخصه‌های پیوسته G در نظر بگیریم، $\widehat{G}$ را می‌توان با عمل $(\xi_1 + \xi_2)(x) = \xi_1(x)\xi_2(x)$ تبدیل به گروهی آبلی کرد. به $\widehat{G}$ گروه دوگان G می‌گویند.

توپولوژی روی $\widehat{G}$ را توپولوژی فشرده‌ـ باز در نظر می‌گیریم. در این صورت می‌توان ثابت کرد که $\widehat{G}$ با توپولوژی فشرده‌ـ باز یک گروه آبلی موضعاً فشرده است.

برای مثال مشخصه‌های پیوسته روی $\mathbb{Z}$ توابعی به شکل $\xi(n) = e^{2\pi in\xi}$ هستند. پس $\widehat{\mathbb{Z}} \cong S^1$. همچنین توابع مشخصه‌ی S^1 به شکل $\xi(x) = x^n$ هستند. در نتیجه $\widehat{S^1} \cong \mathbb{Z}$. در جدول ۱ گروه دوگان برای نمونه‌ای از گروه‌ها آمده است که با چهار نوع تبدیل فوریه‌ی ذکرشده در مقدمه مربوطند.

G	$\widehat{G}$
$\mathbb{R}$	$\mathbb{R}$
S^1	$\mathbb{Z}$
$\mathbb{Z}$	S^1
$\mathbb{Z}/m$	$\mathbb{Z}/m$

جدول ۱: گروه دوگان

حقیقت آن است که اگر $f : G \to \mathbb{C}$، تبدیل فوریه‌ی آن $\hat{f} : \widehat{G} \to \mathbb{C}$ می‌باشد. برای همین جدول ۱ گویای انواع تبدیل فوریه است.

حال به ادامه‌ی بحث گروه‌های دوگان می‌پردازیم. تابع $(.\,,.) : G \times \widehat{G} \to S^1$ را به شکل

$$(x, \xi) = \xi(x) \qquad (۴۲)$$

[13] Haar theorem

[14] Haar measure

تعریف می‌کنیم. این تابع دارای خواص زیر است:

$$(x+x',\xi)=(x,\xi)(x',\xi) \quad (x,\xi+\xi')=(x,\xi)(x,\xi')$$
$$(x,0)=(0,\xi)=1 \quad (-x,\xi)=(x,-\xi)=\overline{(x,\xi)}=(x,\xi)^{-1} \tag{۴۳}$$

می‌توان ثابت کرد که تابع $(.,.)$ پیوسته است. این تابع شهود خوبی از دوگانگی به ما می‌دهد. برای هر $x\in G$، تابع $\hat{x}(\xi)=(x,\xi)$ یک مشخصه برای $\widehat{G}$ می‌باشد. در نتیجه گروه $\widehat{G}$ به طور طبیعی در $\widehat{\widehat{G}}$ می‌نشیند. حال این سؤال پیش می‌آید که آیا $\widehat{\widehat{G}}=G$؟

قضیه ۱۴ (دوگانگی پونتریاگین[15]). *اگر G یک گروه آبلی موضعاً فشرده باشد، آنگاه نگاشت طبیعی $\hat{}:G\to\widehat{\widehat{G}}$ یکریختی و همسانریختی است. در نتیجه به طور طبیعی $G\cong\widehat{\widehat{G}}$*

ناگفته نماند که بسیاری ادعاها از قضایای این بخش از خواص تبدیل فوریه نتیجه می‌شود. لیکن چون مجال پرداختن به آن نیست به طور کلی در این نوشته آورده شده است.

۳.۴ پیچش

در این قسمت فرض کنید G یک گروه آبلی موضعاً فشرده و σـ همبند باشد و m اندازه‌ی هار آن باشد. همچنین در این قسمت $dm(x)$ را با dx نمایش می‌دهیم

تعریف ۱۱ (پیچش[16]). فرض کنید $f,g:G\to\mathbb{C}$ دو تابع باشند. در این صورت پیچش f و g را به شکل

$$f*g(x)=\int_G f(x-y)g(y)dy \tag{۴۴}$$

برای xهایی که

$$\int_G |f(x-y)g(y)|\,dy<\infty \tag{۴۵}$$

تعریف می‌کنیم.

به راحتی می‌توان چک کرد که پیچش دو تابع در صورت وجود، دارای خاصیت جابه‌جایی، شرکت‌پذیری و پخشی در جمع می‌باشد. قضیه‌ی زیر به قسمتی از خواص پیچش اشاره می‌کند.

قضیه ۱۵.

(آ) *اگر $f,g\in L^1(G)$، در این صورت* (۴۵) *برای تقریباً همه‌ی xها صادق است و*

$$\|f*g\|_1\le\|f\|_1\|g\|_1 \tag{۴۶}$$

در نتیجه $L^1(G)$ با اعمال $+$ و $*$ و با نرم $\|\ \|_1$ یک جبر باناخ است.

(ب) *اگر $f\in L^1(G)$ و $g\in L^\infty(G)$ در این صورت $f*g$ کراندار و پیوسته‌ی یکنواخت است.*

(ج) *اگر $f,g\in C_c(X,\mathbb{C})$، در این صورت $f*g\in C_c(X,\mathbb{C})$.*

(د) *اگر $1<p<\infty$ و $\frac{1}{p}+\frac{1}{q}=1$، آنگاه اگر $f\in L^p(G)$ و $g\in L^q(G)$ داریم $f*g\in C_0(G,\mathbb{C})$.*

اثبات. برای قسمت (آ) داریم

$$\begin{aligned}\int_G\int_G|f(x-y)g(y)|\,dydx&=\int_G\left(\int_G|f(x-y)|\,dx\right)|g(y)|\,dy\\&=\int_G\|f\|_1|g(y)|\,dy=\|f\|_1\|g\|_1<\infty\end{aligned} \tag{۴۷}$$

پس طبق قضیه‌ی ۲۱ (آ)

$$\int_G|f(x-y)g(y)|\,dy<\infty \quad a.e.\ x \tag{۴۸}$$

و

$$\|f*g\|_1\le\int_G\int_G|f(x-y)g(y)|\,dydx=\|f\|_1\|g\|_1. \tag{۴۹}$$

برای (ب) واضح است که

$$|f*g|\le\int_G|f(x-y)g(y)|\,dy\le\|f\|_1\|g\|_\infty \tag{۵۰}$$

[15] Pontryagin duality theorem

[16] convolution

پس $f*g$ کران‌دار است. اما برای پیوستگی یکنواخت داریم

$$\begin{aligned}&|f*g(x)-f*g(z)|\\&\leq\int_G|f(x-y)-f(x-z)||g(y)|\,dy\\&\leq\int_G|f(x+y)-f(z+y)|\,dy\;\|g\|_\infty\\&=\|f_{-x}-f_{-z}\|_1\|g\|_\infty\end{aligned}\quad(۵۱)$$

در نتیجه طبق قضیه‌ی ۱۳ عبارت سمت راست را می‌توان به اندازه‌ی دلخواه کوچک کرد.

در قسمت (ج) اولا از قسمت (ب) نتیجه می‌شود که $f*g$ موجود و پیوسته است. حال اگر محمل f برابر با A و محمل g برابر با B باشد، در این صورت

$$f*g(x)=\int_G f(x-y)g(y)dy=\int_B f(x-y)g(y)dy\quad(۵۲)$$

از طرفی دیگر اگر برای هر $y\in B$، $x-y\notin A$ آنگاه انتگرال سمت راست صفر است. یعنی محمل $f*g$ زیرمجموعه‌ی $A+B$ است که به راحتی می‌توان دید که فشرده است.

برای اثبات قسمت (د) ابتدا در نظر بگیرید که $C_c(X,\mathbb{C})$ در $L^p(G)$ و $L^q(G)$ چگال است. لذا می‌توان دنباله‌های $\{f_n\}$ و $\{g_n\}$ از اعضای $C_c(X,\mathbb{C})$ را طوری در نظر گرفت که $\|f_n-f\|_p\to 0$ و $\|g_n-g\|_q\to 0$. در نتیجه طبق نامساوی هلدر (قضیه‌ی ۳)

$$\begin{aligned}&|f_n*g_n(x)-f*g(x)|=|(f_n-f)*(g_n-g)(x)|\\&\quad+|(f_n-f)*g(x)|\;+|f*(g_n-g)(x)|\\&\leq\int_G|f_n(x-y)-f(x-y)||g_n(y)-g(y)|\,dy\\&\quad+\int_G|f_n(x-y)-f(x)||g(y)|\,dy\\&\quad+\int_G|f(x-y)||g_n(y)-g(y)|\,dy\\&\leq\|f_n-f\|_p\|g_n-g\|_q+\|f_n-f\|_p\|g\|_q+\|f\|_p\|g_n-g\|_q\end{aligned}\quad(۵۳)$$

معادله‌ی بالا به ما می‌گوید که $\|f_n*g_n-f*g\|_\infty\to 0$. در نتیجه چون $f_n*g_n\in C_c(X,\mathbb{C})$ (قسمت (ج)) پس $f*g\in C_0(X,\mathbb{C})$. □

می‌توان تعریف پیچش را به اندازه‌های برل روی G تعمیم داد. فرض کنید $M(G)$ مجموعه‌ی همه‌ی اندازه‌های مختلط برل روی G باشد. دقت کنید که برای اندازه‌های مختلط $\|\mu\|<\infty$.

تعریف ۱۲. فرض کنید $\mu,\nu\in M(G)$، در این صورت اندازه‌ی مختلط $\mu*\nu$ به شکل زیر تعریف می‌شود:

$$\mu*\nu(E)=\mu\times\nu\{(x,y)\in G\times G\mid x+y\in E\}\quad(۵۴)$$

به راحتی می‌توان نشان داد که $\mu*\nu\in M(G)$ و رابطه‌ی

$$\int_G f d(\mu*\nu)=\int_G\int_G f(x+y)d\mu(x)d\nu(y)\quad(۵۵)$$

برقرار است. همچنین خواص خوبی مانند شرکت‌پذیری، جابه‌جایی و پخشی در جمع برای عمل پیچش روی اندازه‌ها صادق است و در نتیجه $M(G)$ با عمل جمع و پیچش یک $\mathbb{C}$ـجبر است. به علاوه می‌توان اثبات کرد که

$$\|\mu*\nu\|\leq\|\mu\|\|\nu\|\quad(۵۶)$$

در نتیجه

قضیه ۱۶. *$M(G)$ با عمل جمع و پیچش و نرم تغییرات کلی یک جبر باناخ است.*

همچنین این تعریف تعمیمی از پیچش روی توابع است. زیرا اگر $f\in L^1(G)$ آنگاه $fdx\in M(G)$ و رابطه‌ی

$$(fdx)*(gdx)=(f*g)dx\quad(۵۷)$$

برقرار است.

۵ تبدیل فوریه

در این جا به تبدیل تعریف فوریه می‌پردازیم. در این قسمت فرض کنید G یک گروه آبلی موضعاً فشرده و σـهمبند باشد و اندازه‌ی هار آن را با $dm(x)=dx$ نمایش می‌دهیم. همچنین $\widehat{G}$ را گروه دوگان G می‌گیریم و اندازه‌ی هار آن را با $dm'(\xi)=d\xi$ نمایش می‌دهیم. فعلا اندازه‌ی هار دو گروه را مستقل از هم انتخاب می‌کنیم. در حالی که در بخش ۲.۵ اندازه‌ی هار $\widehat{G}$ را برحسب اندازه‌ی هار G تعیین می‌کنیم.

تعریف ۱۳ (تبدیل فوریه). فرض کنید $f \in L^1(G)$، در این صورت تبدیل فوریه f، تابع $\hat{f} : \widehat{G} \to \mathbb{C}$ است که به شکل

$$\hat{f}(\xi) = \int_G (-x, \xi) f(x) dx \qquad (۵۸)$$

تعریف می‌شود. تبدیل فوریه را با $\mathscr{F}$ و برد آن را با $A(\widehat{G})$ نشان می‌دهیم.

در قسمت‌های ۲.۵ و ۴.۵ تعمیم‌هایی از تبدیل فوریه ارائه خواهیم داد.

۱.۵ خواص اولیه‌ی تبدیل فوریه

در این قسمت به ارائه‌ی خواص اولیه‌ی تبدیل فوریه می‌پردازیم. اولاً واضح است که $\mathscr{F}$ یک تابع خطی است. خواص دیگر آن در زیر آمده است.

قضیه ۱۷.

(آ) $\widehat{f * g} = \hat{f}\hat{g}$

(ب) *اگر* $\tilde{f}(x) = \overline{f(-x)}$، *در این صورت* $\mathscr{F}(\tilde{f}) = \overline{\mathscr{F} f}$.

(ج) $\widehat{f_{x_0}}(\xi) = (-x_0, \xi)\hat{f}(\xi)$

در نتیجه $A(\widehat{G})$ *یک* $\mathbb{C}$*ـجبر خود الحاق است و* F *یک همریختی از جبر* $L^1(G)$ (*عمل* $+$ *و* $*$) *به* $A(\widehat{G})$ *است.*

اثبات. (آ):

$$\begin{aligned}
\widehat{f * g}(\xi) &= \int_G (-x, \xi) f * g(x) dx \\
&= \int_G \int_G (-x, \xi) f(x - y) g(y) dy dx \\
&= \int_G \left(\int_G (-x, \xi) f(x - y) dx \right) g(y) dy \\
&= \int_G \left(\int_G (-x - y, \xi) f(x) dx \right) g(y) dy \\
&= \left(\int_G (-x, \xi) f(x) dx \right) \left(\int_G (-y, \xi) g(y) dy \right) \\
&= \hat{f}(\xi)\hat{g}(\xi) \qquad (۵۹)
\end{aligned}$$

(ب): فرض کنید $g = \tilde{f}$. در این صورت

$$\hat{g}(\xi) = \int_G \overline{(x, \xi) f(-x)} dx = \overline{\left(\int_G (-x, \xi) f(x) dx \right)} = \overline{\hat{f}(\xi)}. \qquad (۶۰)$$

(ج):

$$\begin{aligned}
\widehat{f_{x_0}}(\xi) &= \int_G (-x, \xi) f(x - x_0) dx \\
&= \int_G (-x - x_0, \xi) f(x) dx \\
&= (-x_0, \xi) \int_G (-x, \xi) f(x) dx \\
&= (-x_0, \xi)\hat{f}(\xi) \qquad (۶۱)
\end{aligned}$$

□

اما در مورد ساختار $A(\widehat{G})$ قضیه‌ی زیر تعیین‌کننده است.

قضیه ۱۸.

(آ) $A(\widehat{G}) \subseteq C_0(\widehat{G}, \mathbb{C})$.

(ب) $A(\widehat{G})$ *یک جبر خودالحاق است که هیچ جا صفر نمی‌شود و نقاط* $\widehat{G}$ *را جدا می‌کند. در نتیجه طبق قضیه‌ی استون‌ـوایرشتراس (قضیه‌ی ۱) در* $C_0(\widehat{G}, \mathbb{C})$ *چگال است.*

(ج) *برای هر* $f \in L^1(G)$، $\|\hat{f}\|_\infty \leq \|f\|_1$. *در نتیجه* $\mathscr{F} : L^1(G) \to C_0(\widehat{G}, \mathbb{C})$ *یک تابع پیوسته است.*

اثبات. (آ): اگر $f \in L^1(G)$ و $\xi_n \to \xi$ در این صورت $(-x, \xi_n)f(x) \to (-x, \xi)f(x)$ به صورت نقطه به نقطه و همه‌ی این توابع تحت تسلط $|f|$ هستند. پس طبق قضیه‌ی همگرایی تسلطی (قضیه‌ی ۵) $\hat{f}(\xi_n) \to \hat{f}(\xi)$. پس $\hat{f}$ پیوسته است. اما اثبات این که $\hat{f}$ در بی‌نهایت صفر می‌شود، نیازمند مراجعه به نظریه‌ی گلفاند[۱۷] است که در داخل این نوشته نمی‌گنجد. برای کسب اطلاع از نظریه‌ی گلفاند و ادامه‌ی اثبات به [۵] و [۸] مراجعه شود.

(ب): طبق قضیه‌ی ۱۷ $A(\widehat{G})$ یک جبر خودالحاق است. فرض کنید $K \subseteq G$ یک همسایگی فشرده از همانی باشد. در

[17] Gelfand theory

نتیجه $0 < m(K) < \infty$. برای $\xi_0 \in \widehat{G}$ اگر قرار دهیم $f(x) = \xi_0(x)1_K(x)$ در این صورت $f \in L^1(G)$ و

$$\hat{f}(\xi_0) = \int_G (-x, \xi_0)(x, \xi_0)1_K(x)dx = m(K) > 0 \quad (۶۲)$$

در نتیجه $A(\widehat{G})$ همه جا صفر نمی‌شود. همچنین اگر $\xi_1 \neq \xi_2$ نقطه‌ای مانند x_0 وجود دارد که $\xi_1(x_0) \neq \xi_2(x_0)$ و این نابرابری به خاطر پیوستگی در یک همسایگی از x_0 برقرار است. اگر V را این همسایگی بنامیم و قرار دهیم $f(x) = (\xi_1(x) - \xi_2(x))1_V(x)$، در این صورت نیز $f \in L^1(G)$ و

$$\hat{f}(\xi_1) - \hat{f}(\xi_2) = \int_G \left|\xi_1(x) - \xi_2(x)\right|^2 1_V(x)dx > 0 \quad (۶۳)$$

(به خاطر پیوستگی). در نتیجه $A(\widehat{G})$ نقاط $\widehat{G}$ را جدا می‌کند.

(ج): داریم

(۶۴)

$$|\hat{f}(\xi)| = \left|\int_G (-x, \xi)f(x)dx\right| \leq \int_G |f(x)|dx = \|f\|_1$$

□

تا اینجا ثابت کردیم که تبدیل فوریه $\mathscr{F} : L^1(G) \to C_0(\widehat{G}, \mathbb{C})$ یک همریختی جبری و پیوسته است که تصویر آن در $C_0(\widehat{G}, \mathbb{C})$ چگال است. می‌توان ثابت کرد که این تبدیل یک‌به‌یک است. ولی شمایی از اثبات آن را در قسمت بعد ارائه می‌کنیم.

۲.۵ تبدیل وارون فوریه

در اینجا به تعریف تبدیل وارون فوریه می‌پردازیم. اما در اینجا برعکس قسمت قبل که تبدیل فوریه را روی توابع تعریف کردیم، وارون آن را روی اندازه‌ها تعریف می‌کنیم. دلیل اینکه این تبدیل را وارون فوریه می‌نامیم قضیه‌ی ۲۲ است که به آن اشاره می‌کنیم.

تعریف ۱۴ (تبدیل وارون فوریه[۱۸]). فرض کنید $\mu \in M(\widehat{G})$. در این صورت وارون تبدیل فوریه‌ی آن تابعی مانند $\tilde{\mu} : G \to \mathbb{C}$ است که به شکل

$$\tilde{\mu}(x) = \int_{\widehat{G}} (x, \xi)d\mu(\xi) \quad (۶۵)$$

تعریف می‌شود. تبدیل وارون فوریه را با $\mathscr{F}^{-1}$ و برد آن را با $B(G)$ نشان می‌دهیم.

تعریف تبدیل وارون فوریه بسیار شبیه تعریف تبدیل فوریه است. پس به طور مشابه می‌توان دید که مشابه تمام خواص اشاره‌شده برای تبدیل فوریه (قضایای ۱۷ و ۱۸) در اینجا برقرار است. پس به طور خلاصه

قضیه ۱۹. *تبدیل وارون فوریه $\mathscr{F}^{-1} : M(\widehat{G}) \to C_0(G, \mathbb{C})$ یک $\mathbb{C}$ـ همریختی پیوسته است و برد آن چگال است.*

در اینجا ثابت خواهیم کرد که تبدیل وارون فوریه یک‌به‌یک است. اما نکته‌ی مهمی که وجود دارد این است که می‌توان از دوگانگی G و $\widehat{G}$ استفاده کرد و به طور مشابه ثابت کرد که تبدیل فوریه یک‌به‌یک است.

قضیه ۲۰. *تبدیل وارون فوریه یک‌به‌یک است.*

اثبات. چون $\mathscr{F}^{-1}$ خطی است، کافی است ثابت کنیم هسته‌ی آن صفر است. فرض کنید $\mathscr{F}^{-1}(\mu) = 0$. در این صورت برای هر $f \in L^1(G)$ داریم

$$\int_{\widehat{G}} \hat{f}(\xi)d\mu(\xi) = \int_{\widehat{G}}\int_G (-x, \xi)f(x)dxd\mu(\xi) = \int_G \left(\int_{\widehat{G}} (-x, \xi)d\mu(\xi)\right) f(x)dx = \int_G \tilde{\mu}(-x)f(x)dx = 0 \quad (۶۶)$$

اما از آن جا که $A(\widehat{G})$ در $C_0(\widehat{G}, \mathbb{C})$ چگال است، $\mu = 0$. (رجوع شود به قضیه‌ی ۱۰) □

چون تبدیل وارون فوریه یک‌به‌یک است، وارون دارد و تبدیل $(\mathscr{F}^{-1})^{-1} : B(G) \to M(\widehat{G})$ یک همریختی است. برای اینکه مشخص شود که چرا به $\mathscr{F}^{-1}$ تبدیل وارون فوریه می‌گوییم، باید به رابطه‌ی $(\mathscr{F}^{-1})^{-1}$ و $\mathscr{F}$ پی ببریم. قضیه‌ی وارون فوریه (قضیه‌ی ۲۲) به ما می‌گوید که این دو تبدیل روی $L^1(G) \cap B(G)$ یکی هستند. به این معنا که

$$(\mathscr{F}^{-1})^{-1}(f) = \hat{f}d\xi \qquad f \in L^1(G) \cap B(G) \quad (۶۷)$$

[18] inverse Fourier transform

در نتیجه می‌توان تابع $(\mathscr{F}^{-1})^{-1}$ را به نحوی تعمیم تبدیل فوریه در نظر گرفت. با این تعریف تبدیل فوریه تابع f به این مفهوم است که این تابع چگونه ترکیب خطی‌ای از توابع مشخصه است. با این تعریف از تبدیل فوریه، تبدیل فوریه توابع به شکل $e^{2\pi ix\xi}$ در تبدیل فوریه پیوسته معنا پیدا می‌کند.

در آخر این بخش به مشخص کردن اعضای $B(G)$ می‌پردازیم.

تعریف ۱۵. تابع $\phi : G \to \mathbb{C}$ را مثبت معین می‌نامیم هرگاه برای هر $x_1, ..., x_n \in G$ ماتریس $[\phi(x_i - x_j)]_{i,j}$ مثبت معین باشد. یعنی برای هر $c_1, ..., c_n \in \mathbb{C}$،

$$\sum_{i,j} c_i \bar{c}_j \phi(x_i - x_j) \geq 0. \qquad (۶۸)$$

توابع مثبت معین خواص مهمی دارد، از جمله اینکه

$$\phi(-x) = \overline{\phi(x)}, \qquad \big|\phi(x)\big| \leq \phi(0). \qquad (۶۹)$$

برای اطلاع بیشتر از خواص توابع مثبت معین به [۸] مراجعه کنید. اگر μ یک اندازه‌ی مثبت با $\|\mu\| < \infty$ روی $\widehat{G}$ باشد، در این صورت تابع

$$\phi(x) = \int_{\widehat{G}} (x, \xi) d\mu(\xi) = \tilde{\mu}(x) \qquad (۷۰)$$

مثبت معین است. برای دیدن این مطلب توجه کنید که

$$\begin{aligned} \sum_{n,m} c_n \bar{c}_m \phi(x_n - x_m) &= \int_{\widehat{G}} \left(\sum_{m,n} c_n \bar{c}_m \xi(x_n) \overline{\xi(x_m)} \right) d\mu(\xi) \\ &= \int_{\widehat{G}} \left| \sum_n c_n \xi(x_n) \right|^2 d\mu(\xi) \geq 0. \end{aligned} \qquad (۷۱)$$

اما قضیه‌ی زیر که به قضیه‌ی بوچنر[19] معروف است عکس مطلب بالا را نیز ادعا می‌کند.

قضیه ۲۱ (بوچنر). *تابع $\phi : G \to \mathbb{C}$ مثبت معین است اگر و فقط اگر اندازه‌ی مثبت $\mu \in M(\widehat{G})$ موجود باشد که $\phi = \mathscr{F}^{-1}(\mu)$.*

اثبات قضیه‌ی بالا نیازمند جزئیات بیشتری است که در این نوشته نمی‌گنجد. برای مشاهده‌ی اثبات به [۸] مراجعه شود. نتیجه مهم قضیه‌ی بالا این است که اعضای $B(G)$ به شکل ترکیب خطی‌ای از توابع مثبت معین هستند، زیرا هر اندازه‌ی مختلطی را می‌توان به صورت ترکیب خطی‌ای از اندازه‌های مثبت نوشت. (رجوع شود به قضیه‌ی ۹)

۳.۵ قضیه‌ی وارون فوریه

در بخش قبل توضیحاتی درباره قضیه‌ی وارون فوریه داده شد. در این قسمت می‌خواهیم به بیان دقیق صورت مسئله و اثبات آن بپردازیم. برای شروع کار ابتدا لازم است لمی را ثابت کنیم.

لم ۱.

(آ) *فرض کنید $f \in L^1(G)$ و $\tilde{f}(x) = \overline{f(-x)}$، در این صورت اگر $g = f * \tilde{f}$ آنگاه g مثبت معین است.*

(ب) *اگر $K \subseteq \widehat{G}$ فشرده باشد، در این صورت $g \in L^1(G)$ موجود است که $\hat{g}$ روی K اکیداً مثبت باشد و g مثبت معین باشد.*

اثبات. (آ): داریم

$$\begin{aligned} &\sum_{m,n} c_n \bar{c}_m g(x_n - x_m) \\ &\quad = \sum_{m,n} c_n \bar{c}_m \int_G f(x_n - x_m - y) \overline{f(-x)} dy \\ &\quad = \int_G \sum_{m,n} c_n \bar{c}_m f(x_n - y) \overline{f(x_m - y)} dy \\ &\quad = \int_G \left| \sum_{m,n} c_n f(x_n - y) \right|^2 dy \geq 0 \end{aligned} \qquad (۷۲)$$

(ب): برای هر $\xi \in K$ طبق قضیه‌ی ۲۹ (ب) می‌توان $u_\xi \in L^1(G)$ را طوری پیدا کرد که $\hat{u}_\xi(\xi) \neq 0$. به خاطر پیوستگی، $\hat{u}_\xi$ در یک همسایگی V_ξ از ξ ناصفر است. طبق فشردگی K می‌توان متناهی تا از آن‌ها یعنی $V_1, ..., V_n$ را طوری

[19] Bochner's theorem

انتخاب کرد که K را پوشش دهند. در این صورت اگر قرار دهیم $g = u_1 * \tilde{u}_1 + ... + u_n * \tilde{u}_n$ آنگاه

$$\hat{g} = |\hat{u}_1|^2 + ... + |\hat{u}_n|^2 \qquad (۷۳)$$

در نتیجه $\hat{g}$ روی K اکیداً از صفر بزرگتر است. □

حال به بیان صورت قضیه‌ی وارون فوریه می‌پردازیم.

قضیه ۲۲ (قضیه‌ی وارون فوریه). *اگر $f \in L^1(G) \cap B(G)$ آنگاه $\hat{f} \in L^1(\widehat{G})$ و می‌توان اندازه‌های هار روی G و $\widehat{G}$ را طوری انتخاب کرد که برای هر $f \in L^1(G) \cap B(G)$، اگر $\mu = (\mathscr{F}^{-1})^{-1}(f)$، یعنی*

$$f(x) = \int_{\widehat{G}} (x, \xi) d\mu(\xi) \qquad (۷۴)$$

آنگاه

$$d\mu(\xi) = \hat{f}(\xi) d\xi \qquad (۷۵)$$

اثبات. در این اثبات $(\mathscr{F}^{-1})^{-1}(f)$ را با μ_f و $L^1(G) \cap B(G)$ را با B^1 نشان می‌دهیم. اگر $f \in B^1$ و $h \in L^1(G)$، در این صورت

$$(h * f)(0) = \int_G h(-x) f(x) dx$$
$$= \int_G \int_{\widehat{G}} h(-x)(x, \xi) dx d\mu_f(\xi) = \int_{\widehat{G}} \hat{h}(\xi) d\mu_f(\xi) \qquad (۷۶)$$

در نتیجه اگر $g \in B^1$،

$$\int_{\widehat{G}} \hat{h}(\xi)\hat{g}(\xi) d\mu_f(\xi) = (h * g) * f\ (0)$$
$$= (h * f) *\ g(0) = \int_{\widehat{G}} \hat{h}(\xi)\hat{f}(\xi) d\mu_g(\xi) \qquad (۷۷)$$

پس چون رابطه‌ی (۷۷) برای هر $h \in L^1(G)$ درست است و $A(\widehat{G})$ در $C_0(\widehat{G}, \mathbb{C})$ چگال است، داریم

$$\int_{\widehat{G}} \phi \hat{g} d\mu_f = \int_{\widehat{G}} \phi \hat{f} d\mu_g \qquad \forall \phi \in C_0(\widehat{G}, \mathbb{C}) \qquad (۷۸)$$

حال تابع خطی $T : C_c(\widehat{G}, \mathbb{C}) \to \mathbb{C}$ را به این شکل تعریف می‌کنیم: فرض کنید $\psi \in C_c(\widehat{G}, \mathbb{C})$ و $\operatorname{supp} \psi \subseteq K$ که K فشرده است. در این صورت اگر $g \in B^1$ طوری باشد که $\hat{g}$ روی K اکیداً مثبت باشد. در این صورت تعریف می‌کنیم

$$T(\psi) = \int_{\widehat{G}} \frac{\psi}{\hat{g}} d\mu_g \qquad (۷۹)$$

باید نشان دهیم T خوش‌تعریف است. اولا طبق لم ۱ (ب) چنین gای وجود دارد. ثانیا اگر $f \in B^1$ تابع دیگری باشد که خواص g را دارا باشد، طبق (۷۸) داریم

$$\int_{\widehat{G}} \frac{\psi}{\hat{f}\hat{g}} \hat{f} d\mu_g = \int_{\widehat{G}} \frac{\psi}{\hat{f}\hat{g}} \hat{g} d\mu_f \qquad (۸۰)$$

پس (۷۹) خوش‌تعریف است. واضح است که T خطی است. همچنین طبق لم ۱ (ب) می‌توان در (۷۹) g را طوری انتخاب کرد که مثبت معین باشد. در نتیجه $\mu_g \geq 0$، پس اگر $\psi \geq 0$ آنگاه به وضوح $T(\psi) \geq 0$. پس T یک تابع خطی مثبت است. می‌خواهیم نشان دهیم T تحت انتقال ناوردا است. فرض کنید $f(x) = (-x, \xi_0) g(x)$. در این صورت به راحتی می‌توان دید که $\hat{f}(\xi) = \hat{g}(\xi + \xi_0)$ و $d\mu_f(\xi) = d\mu_g(\xi + \xi_0)$ (رجوع شود به قضیه‌ی ۲۹ (ج)) در این صورت اگر $\psi_0(\xi) = \psi(\xi - \xi_0)$، آنگاه

$$T(\psi_0) = \int_{\widehat{G}} \frac{\psi(\xi - \xi_0)}{\hat{g}(\xi)} d\mu_g(\xi)$$
$$= \int_{\widehat{G}} \frac{\psi(\xi)}{\hat{g}(\xi + \xi_0)} d\mu_g(\xi + \xi_0)$$
$$= \int_{\widehat{G}} \frac{\psi(\xi)}{\hat{f}(\xi)} d\mu_f(\xi) = T(\psi) \qquad (۸۱)$$

پس T تحت انتقال راست ناوردا است. در نتیجه اندازه‌ی هار روی $\widehat{G}$ موجود است که

$$T(\psi) = \int_{\widehat{G}} \psi d\xi \qquad \forall \psi \in C_c(\widehat{G}, \mathbb{C}) \qquad (۸۲)$$

(رجوع شود به قضایای ۱۱ و ۱۲) حال اگر $f \in B^1$، برای هر $\psi \in C_c(\widehat{G}, \mathbb{C})$ داریم

$$\int_{\widehat{G}} \psi \hat{f} d\xi = T(\psi \hat{f}) = \int_{\widehat{G}} \frac{\psi}{\hat{g}} \hat{f} d\mu_g = \int_{\widehat{G}} \psi d\mu_f \qquad (۸۳)$$

پس چون رابطه‌ی (۸۳) برای هر $\psi \in C_c(\widehat{G}, \mathbb{C})$ درست است، پس روی هر زیرمجموعه‌ی فشرده‌ی $\widehat{G}$، $\hat{f} d\xi = d\mu_f$. در نتیجه طبق (۳۲) روی زیرمجموعه‌های فشرده‌ی $\widehat{G}$ رابطه‌ی

$$|d\mu_f| = |\hat{f}| d\xi \qquad (۸۴)$$

برقرار است. حال اگر K_n دنباله‌ای از زیرمجموعه‌های فشرده‌ی $\widehat{G}$ باشند که به طور صعودی به $\widehat{G}$ همگرا است، آنگاه طبق قضیه‌ی همگرایی یکنوا (قضیه‌ی ۴) داریم

$$\int_{\widehat{G}} |\hat{f}|d\xi = \lim_{n\to\infty} \int_{\widehat{G}} 1_{K_n}|\hat{f}|d\xi$$
$$= \lim_{n\to\infty} |\mu_f|(K_n) = |\mu_f|(\widehat{G}) < \infty \quad (۸۵)$$

در نتیجه $\hat{f} \in L^1(\widehat{G})$ و

$$d\mu_f = \hat{f}d\xi \quad (۸۶)$$

□

از قضیه‌ی وارون فوریه می‌توان این استفاده را کرد که اندازه‌ی هار $\widehat{G}$ را بر حسب اندازه‌ی هار G بهنجار کرد.

۴.۵ قضیه‌ی پلانچرل

در این قسمت می‌خواهیم تبدیل فوریه را به $L^2(G)$ تعمیم دهیم. درواقع قضیه‌ی زیر می‌گوید تبدیل فوریه را می‌توان به یک ایزومتری بین $L^2(G)$ و $L^2(\widehat{G})$ تعمیم داد.

قضیه ۲۳ (پلانچرل[۲۰]). *برای هر $f \in L^1(G) \cap L^2(G)$، داریم*

$$\|\hat{f}\|_2 = \|f\|_2 \quad (۸۷)$$

و تصویر $L^1(G) \cap L^2(G)$ تحت $\mathscr{F}$ در $L^2(\widehat{G})$ چگال است. در نتیجه $\mathscr{F}$ را می‌توان به یک ایزومتری بین $L^2(G)$ و $L^2(\widehat{G})$ تعمیم داد.

اثبات. فرض کنید $f \in L^1(G) \cap L^2(G)$. اگر $\tilde{f}(x) = \overline{f(-x)}$ و $g = f * \tilde{f}$، آنگاه طبق لم ۱ (آ) g مثبت معین است. پس طبق قضیه‌ی بوچنر $g \in B(G)$، در نتیجه طبق قضیه‌ی وارون فوریه

$$\int_{\widehat{G}} |\hat{f}|^2 d\xi = \int_{\widehat{G}} \hat{g}d\xi = g(0)$$
$$= \int_G f(x)\tilde{f}(-x)dx = \int_G |f|^2 dx. \quad (۸۸)$$

فرض کنید Φ تصویر $L^1(G) \cap L^2(G)$ باشد. برای اینکه ثابت کنیم Φ در $L^2(\widehat{G})$ چگال است، کافی است ثابت کنیم $\Phi^\perp = 0$ فرض کنید $\psi \in \Phi^\perp$، یعنی برای هر $\phi \in \Phi$، $\int_{\widehat{G}} \phi\bar{\psi}d\xi = 0$. چون $L^1(G) \cap L^2(G)$ تحت انتقال ناوردا است، پس Φ تحت ضرب در (x, ξ) ناوردا است. در نتیجه برای هر $x \in G$،

$$\int_{\widehat{G}} \phi(\xi)\overline{\psi(\xi)}(x, \xi)d\xi = 0 \quad (۸۹)$$

در نتیجه طبق قضیه‌ی ۲۰، $\phi\bar{\psi} = 0$ $a.e.$. از طرفی می‌توان مانند قضیه‌ی ۲۹ (ب) نشان داد که Φ هیچ جا صفر نمی‌شود. در نتیجه $\psi = 0$ $a.e.$ □

تعمیم تبدیل فوریه به یک تناظر یک‌به‌یک بین $L^2(G)$ و $L^2(\widehat{G})$ در بسیاری از زمینه‌ها، بخصوص نظریه‌ی کوانتوم کاربرد دارد. در نظریه‌ی گروه‌های کوانتومی، توابع موج مکانی شرودینگر، اعضای فضای $L^2(G)$ هستند که گروه G بیانگر مکان است و توابع موج تکانه‌ای اعضای $L^2(\widehat{G})$ هستند که اعضای $\widehat{G}$ نشان‌دهنده‌ی تکانه‌های مختلف هستند و برای تابع مکان ψ، $\hat{\psi}$ تابع موج تکانه‌ی آن است. درواقع هر عضو $\widehat{G}$ بیانگر تابع موج مکانی ذره‌ای با تکانه‌ی ثابت است و تابع موج مکانی، ترکیب خطی‌ای از اینها باید باشد، که این تعریف به L^2 تعمیم پیدا می‌کند. برای اطلاعات بیشتر رجوع شود به [۶].

مراجع

[1] Kenneth A. Ross Edwin Hewitt. Abstract Harmonic Analysis. Springer, 1963.

[2] Karl Stomberg Edwin Hewwit. Real and Abstract Analysis. Springer-Verlag, 1965.

[3] Revaz Gamkrelidze. Topological Groups. Taylor & Francis, 1987.

[4] Patrick Fitzpatrick Halsey Royden. Real analysis. Pearson Education, Inc, 4th edition, 2010.

[20] Plancherel

[5] Eberhard Kaniuth. A Course in Commutative Banach Algebra. Springer New York, 2008.

[6] Christian Kassel. Quantum Groups. Springer-Verlag, 1995.

[7] Inder K. Rana. An Introduction to Measure and Integration. American Mathematical Society, 2002.

[8] Walter Rudin. Fourier Analysis on Groups. Interscience publishers, 1st edition, 1962.

قدم زن تصادفی ساده و شبکه‌های الکتریکی

سایه خانی‌ها

چکیده

ارتباط جالب و عمیقی بین قدم زن تصادفی ساده روی گراف‌ها و تحلیل شبکه‌های الکتریکی وجود دارد. مقادیری مانند زمان جابه‌جایی و احتمال فرار به وسیله روش‌های موجود در نظریه مدارها قابل محاسبه هستند. همچنین از مفهوم مقاومت معادل می‌توانیم برای تشخیص گذرا یا بازگشتی بودن یک قدم زن تصادفی ساده روی گراف‌ها استفاده کنیم.

۱ قدم زن تصادفی ساده

یک گراف همبند $G = (V, E)$ در نظر بگیرید، یک قدم زدن تصادفی ساده روی G از یک رأس مشخص در گراف شروع می‌شود و قدم زن تصادفی در هر حرکت، از یک رأس به رأس مجاوری که به احتمال یکنواخت از بین رئوس مجاور انتخاب شده، حرکت می‌کند. به عنوان مثال می‌توانید قدم زدن تصادفی را روی گراف زیر در نظر بگیرید،

شکل ۱: گراف مسیر

فرض کنید از یک نقطه مثلاً نقطه k قدم زدن را شروع کنیم. می‌توانیم به این قدم زدن تصادفی به عنوان مسئله ورشکستگی قمار باز نگاه کنیم. قماربازی با k دلار شروع به بازی می‌کند، بازی این چنین است که او در هر مرحله سکه‌ای سالم را پرتاپ می‌کند، اگر سکه شیر بیاید، قمارباز یک دلار می‌برد و اگر خط بیاید، یک دلار از دست می‌دهد. در این صورت مایلیم جواب این سؤال را پیدا کنیم که احتمال رسیدن قمارباز به مبلغ N دلار قبل از ورشکستگی، یعنی رسیدن به مبلغ 0 دلار، چقدر است؟

برای پاسخ دادن به این سؤال تابع $p(k)$ را احتمال این که قمار باز با شروع از k دلار قبل از ورشکستگی به مبلغ N دلار برسد، تعریف می‌کنیم. واضح است که $p(0) = 0$ و $p(N) = 1$. همچنین اگر $k \notin \{0, N\}$، در این صورت می‌توانیم با شرطی کردن روی قدم بعد رابطه ای بازگشتی به شکل زیر بنویسیم،

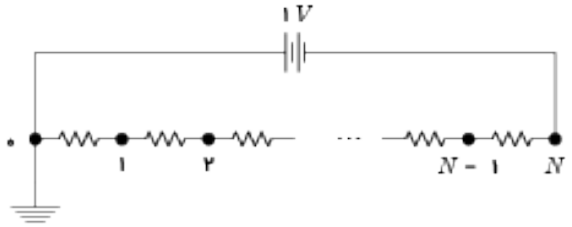
$$p(k) = \Pr(A) \times p(k+1) + \Pr(B) \times p(k-1)$$

در رابطه بالا A این پیشامد است که قمارباز در پرتاب بعدی ۱ دلار برنده شود و B این پیشامد که قمارباز در پرتاب بعدی ۱ دلار ببازد.

اگر رابطه بازگشتی بالا را با شرایط اولیه‌ای که داریم حل کنیم، جوابی به صورت $p(k) = \frac{k}{N}$ به دست می‌آید. در ادامه خواهیم دید که خیلی راحت‌تر می‌توانیم به این جواب برسیم.

۲ شبکه‌های الکتریکی

بیایید به این سؤال به ظاهر نامربوط به بحث قبل در مورد شبکه‌های الکتریکی پاسخ دهیم؛ مدار زیر را که در آن هر مقاومت ۱ اهم است، در نظر بگیرید،



فرض کنید $f(k)$ اختلاف پتانسیل بین نقطه k و نقطه 0 باشد، در این صورت مقدار $f(k)$ چقدر است؟

برای پاسخ دادن به این سؤال ابتدا نیاز به مرور چند نماد داریم. برای هر دو رأس $a, b \in V$ اختلاف پتانسیل آن دو راس را با $\phi(a,b)$ نشان می‌دهیم. اگر $\{a,b\} \in E$، آنگاه $I(a,b)$ نشان‌دهنده جریان الکتریکی جاری از نقطه‌ی a تا b و $R_{a,b}$ مقاومت یال (a,b) است. همچنین حقایق آشنای زیر را درباره مدارهای الکتریکی داریم،

قانون اهم: در یال جهت‌دار (a,b) داریم $\phi(a,b) = I(a,b)R_{a,b}$.

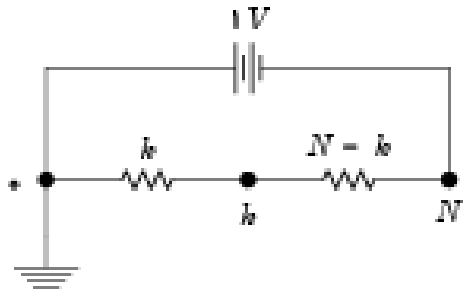
قانون اول کرشهف: در هر رأس v از مدار، جمع جریان‌هایی که به آن رأس وارد می‌شود با جمع جریان‌هایی که از آن خارج می‌شود برابر است، یعنی $\sum_{u:u\sim v} I(u,v) = 0$.

قانون دوم کرشهف: برای هر v_0، v_1، ... و $v_k = v_0$ داریم $\sum \phi(v_i, v_{i+1}) = 0$. یعنی در هر حلقه یا مسیر بسته مجموع جبری اختلاف پتانسیل در المان‌های مدار برابر صفر است.

حال می‌توانیم به سؤالمان درباره $f(k)$ برگردیم. می‌توان دید که $f(0) = 0$ و $f(N) = 1$. همچنین اگر $k \notin \{0, N\}$ داریم،

$$\begin{aligned} 0 &= I(k-1,k) + I(k+1,k) \\ &= \phi(k-1,k) + \phi(k+1,k) \\ &= \phi(k-1,0) - \phi(k,0) \\ &\quad + \phi(k+1,0) - \phi(k,0) \\ &= (f(k-1) - f(k)) + (f(k+1) - f(k)) \\ &= f(k+1) + f(k-1) - 2f(k) \end{aligned}$$

با توجه به روابط بالا داریم $f(k) = \frac{f(k+1)+f(k-1)}{2}$. با برگشتن به قدم زن تصادفی ساده‌ای که داشتیم، می‌بینیم که توابع f و p هر دو در شرایط مرزی و روابط بازگشتی یکسانی صدق می‌کنند. این مطلب نشان می‌دهد که این دو باید با هم برابر باشند. در واقع به سادگی می توان این معادله ی تفاضلی را حل کرد و جواب آن تحت این شرط مرزی یکتاست. گاهی اوقات مراجعه به نسخه مداری مسئله و تحلیل مدار می‌تواند حل آن را راحت‌تر کند. از آنجایی که مقاومت‌های بین نقطه 0 تا k در مدار فوق به صورت سری بسته شده‌اند، می‌توانیم آن‌ها را با یک مقاومت بزرگتر جایگزین کنیم. همچنین همین کار را می‌توانیم با مقاومت‌های بین نقطه k تا N نیز انجام دهیم. در نتیجه مدار معادلی به شکل زیر خواهیم داشت،



از آنجایی که جریان در کل مدار برابر با $\frac{1}{N}$ است، پس می‌بینیم که $p(k) = f(k) = \phi(k,0) = \frac{k}{N}$. یعنی در این حالت برای پیدا کردن $p(k)$ راحت‌تر بود که به جای نگاه کردن به قدم زن تصادفی ساده به مدار نگاه کنیم. برای هر گراف دیگری به جز گراف مسیر

نیز می‌توانیم استدلالی مشابه آنچه گفته شد، انجام دهیم.

قضیه ۱. *گراف متناهی و همبند $G=(V,E)$ و دو رأس a و z از آن را در نظر بگیرید. در قدم زدن تصادفی با شروع از نقطه k، احتمال برخورد به رأس a قبل از برخورد به رأس z برابر است با اختلاف پتانسیل بین نقاط k و z در شبکه‌ای مشابه که هر یال آن دارای مقاومت 1 است و بین نقاط a و z اختلاف پتانسیل 1 اعمال شده است.*

این قضیه در واقع به یکتایی توابع هارمونیک روی گراف‌های متناهی مربوط می‌شود. یک تابع $f:V\to\mathbb{R}$ ، یک تابع هارمونیک در رأس v است، اگر

$$f(v)=\frac{1}{\deg(v)}\sum_{u\sim v}f(u)$$

لم ۱. *فرض کنید $G=(V,E)$ یک گراف متناهی همبند و $B\subset V$ ناتهی باشد. اگر f روی $V\backslash B$ هارمونیک و روی B مساوی صفر باشد، در این صورت f متحد با صفر خواهد بود.*

اثبات. *ابتدا نشان می‌دهیم که $f\leq 0$. از آن‌جایی که گراف متناهی است، رأسی مانند v_0 در آن وجود دارد که $f(v_0)=M$ و M بیشینه مقدار f روی V است. حال فرض کنید $w\sim v_0$. در این صورت اگر $f(w)<f(v_0)$ داریم،*

$$\begin{aligned}f(v_0)&=\frac{1}{\deg(v_0)}\sum_{u\sim v_0}f(u)\\&=\frac{1}{\deg(v_0)}.f(w)\\&\quad+\frac{1}{\deg(v_0)}\sum_{u\sim v_0,u\neq w}f(u)<M\end{aligned}$$

که این ممکن نیست، بنابراین $f(w)=f(v_0)$. از آن‌جایی که گراف همبند است، مسیری از v_0 به رأسی از B مثل y وجود دارد. با تکرار استدلال بالا نتیجه می‌گیریم که $f(v_0)=f(y)=0$، بنابراین بیشینه مقدار f روی V مساوی صفر است. با استدلالی مشابه روی $-f$ نتیجه می‌گیریم که $f\geq 0$ و این اثبات را کامل می‌کند.

اثبات قضیه ۱. *اگر تابع مربوط به قدم زن تصادفی را با p و تابع مربوط به نظریه مدارها را با f نمایش دهیم، در هر رأس v به جز a و z داریم،*

$$\begin{aligned}p(v)&=\sum_{u\sim v}\Pr(\text{اولین گام از } v \text{ به } u \text{ باشد})\times p(u)\\&=\frac{1}{\deg(v)}\sum_{u\sim v}p(u)\end{aligned}$$

بنابراین p روی $V\backslash\{a,z\}$ هارمونیک است. به طور مشابه با استفاده از قوانین کرشهف داریم،

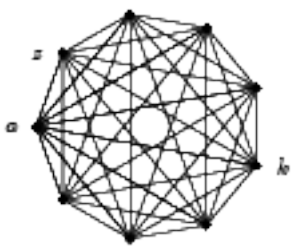
$$\begin{aligned}0&=\sum_{u\sim v}I(u,v)\\&=\sum_{u\sim v}\phi(u,v)\\&=\sum_{u\sim v}\phi(u,z)-\phi(v,z)\\&=\Big(\sum_{u\sim v}f(u)\Big)-\deg(v)f(v).\end{aligned}$$

رابطه بالا نشان می‌دهد که f نیز یک تابع هارمونیک روی $V\backslash\{a,z\}$ است. در نتیجه می‌توانیم مشاهده کنیم که

$$(f-p)(a)=(f-p)(z)=0$$

و $f-p$ روی $V\backslash\{a,z\}$ هارمونیک است. با استفاده از لم ۱ نتیجه می‌شود که $f-p\equiv 0$ یعنی $f=p$.

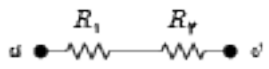
از این رابطه دو طرفه بین قدم زدن تصادفی روی گراف‌ها و شبکه‌های الکتریکی می‌توانیم برای ساده‌تر کردن مسائل احتمالاتی و همچنین سؤالات مربوط به تحلیل مدارها استفاده کنیم. مثال ورشکستگی قمارباز مثالی بود که در آن استفاده از تحلیل مداری ساده‌تر از نگاه احتمالاتی بود ولی ممکن است برعکس این اتفاق نیز بیفتد. برای مثال مداری با n گره در نظر بگیرید که در آن هر رأس به همه رئوس دیگر متصل است یعنی گراف کامل با n رأس را در نظر بگیرید،



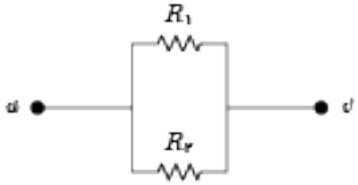
اگر ما دو رأس a و z را در نظر بگیریم و اختلاف پتانسیل 1 بین آن‌ها ایجاد کنیم، اختلاف پتانسیل بین یک رأس دیگر و z چقدر خواهد بود؟ به سؤال متناظر با این سؤال در قدم زدن تصادفی ساده نگاه کنید، می‌بینید که در قدم زدن تصادفی ساده احتمال رسیدن به دو رأس a و z با هم برابر است پس برای همه k هایی که a و z نیستند، داریم $p(k) = \frac{1}{2}$. این یعنی اختلاف پتانسیل هر رأس به جز a با z برابر با $\frac{1}{2}$ است.

۳ زمان جابه‌جایی و مقاومت معادل

به یاد دارید که ما در مثال ورشکستگی قمارباز k مقاومت کوچک را با یک مقاومت بزرگ جایگزین کردیم. این مثالی از به دست آوردن مقاومت معادل است. مقاومت معادل بین دو نقطه u و v میزان اختلاف پتانسیل مورد نیاز بین آن دو نقطه است که بتواند یک واحد جریان را از نقطه u به نقطه v بفرستد. این کمیت را با $R(u,v)$ نشان می دهیم. یادآوری می‌کنیم که اگر دو مقاومت سری باشند مانند شکل زیر،



در این صورت $R(u,v) = R_1 + R_2$. اگر دو مقاومت موازی باشند،



در این صورت $R(u,v) = \frac{1}{\frac{1}{R_1}+\frac{1}{R_2}}$. یک بیان احتمالاتی از مقاومت معادل بر اساس احتمال فرار وجود دارد.

گزاره ۱. *فرض کنید که $p(a,z)$ این احتمال باشد که قدم زن تصادفی ساده با شروع از a قبل از بازگشت به a به z برخورد کند. در این صورت،*

$$p(a,z) = \frac{1}{\deg(a)R(a,z)}$$

با توجه به گزاره بالا می‌بینید که مفهوم مقاومت معادل برای پی بردن به این مطلب که رسیدن از یک رأس به رأس دیگر چقدر دشوار است، مفید است.

$H_{u,v}$ را امید ریاضی زمان برخورد قدم زن تصادفی ساده به راس v با شروع حرکت از راس u تعریف کنید. توجه کنید که $H_{u,v}$ لزوماً با $H_{v,u}$ برابر نیست. برای مثال در گراف زیر،

قدم زن تصادفی که از u شروع به حرکت می‌کند، همواره در اولین قدم به v می‌رسد. پس $H_{u,v} = 1$. ولی $H_{v,u} > 1$ زیرا یک قدم زن تصادفی که از v شروع به حرکت کرده است، ممکن است که اول به سمت راست حرکت کند. برای این که مشکل عدم تقارن را رفع کنیم، زمان جابه‌جایی بین دو رأس u و v را به صورت $C_{u,v} = H_{u,v} + H_{v,u}$ تعریف می‌کنیم. به طرز شگفت انگیزی، این زمان جابه‌جایی رابطه‌ی نزدیکی با مقاومت معادل دارد.

گزاره ۲. *برای هر دو رأس u و v در گراف همبند $G = (V,E)$ داریم،*

$$C_{u,v} = 2|E|R(u,v)$$

این گزاره نیز یک رابطه دوطرفه بین قدم زدن تصادفی ساده و تحلیل مدارها به دست می‌دهد. بیایید این رابطه را روی گراف مسیر (شکل ۱) نگاه کنیم. $H_{0,N}$ در این مثال چقدر است؟ چون این گراف متقارن است، پس اینجا $H_{0,N} = H_{N,0}$. در نتیجه با استفاده از گزاره قبل خواهیم داشت،

$$H_{0,N} = |E|R(0,N) = N.N = N^2$$

همان طور که می‌بینید این رابطه بدون ذره‌ای تلاش متوسط زمان برخورد را به ما می‌دهد. درصورتی که اگر می‌خواستیم به مسئله از دیدگاه احتمالاتی نگاه کنیم اوضاع پیچیده‌تر بود. این رابطه می‌تواند

برای حل مسائل مربوط به مدارها هم مفید باشد. فرض کنید می خواهیم در گراف کامل n رأسی مقاومت معادل بین دو رأس a و b را پیدا کنیم. با استفاده از گزاره ۲ می‌توانیم بنویسیم،

$$R(a,b) = \frac{H_{a,b} + H_{b,a}}{2|E|} = \frac{H_{a,b}}{\binom{n}{2}}$$

می‌توانیم به راحتی $H_{a,b}$ را محاسبه کنیم. قدم زن تصادفی هرجایی به جز نقطه b باشد، احتمال رفتن آن به نقطه b در قدم بعد برابر $\frac{1}{n-1}$ است. این نشان می‌دهد که تعداد قدم‌های لازم برای رسیدن از a به b یک متغیر تصادفی هندسی با پارامتر $\frac{1}{n-1}$ است. پس امید ریاضی این متغیر هندسی $n-1$ است. با جای گذاری این مقدار در رابطه قبل خواهیم داشت،

$$R(a,b) = \frac{n-1}{\binom{n}{2}} = \frac{2}{n}.$$

۴ کران بالایی برای زمان پوشش

برای گراف متناهی و همبند G و هر رأس v از آن، زمان پوشش [1] گراف با شروع از رأس v را با C_v نمایش می‌دهیم. C_v در واقع متوسط تعداد قدم‌های لازم برای مشاهده‌ی همه رأس‌ها است، با این فرض که قدم زدن تصادفی را از رأس v شروع کرده‌باشیم. زمان پوشش گراف را $C(G) := \max_{v\in V} C_v$ تعریف می‌کنیم. برای گراف خطی به طول N، زمان پوشش با شروع از نقطه‌ی صفر، در واقع همان زمان برخورد به نقطه‌ی N است که قبلاً مقدار آن را برابر با $C_0 = N^2$ حساب کردیم.

یک سؤال که مطرح می‌شود این است که آیا ارتباطی بین زمان پوشش و تعداد رأس‌ها و یال‌های گراف وجود دارد؟

قضیه ۲. *برای هر گراف همبند G داریم*

$$C(G) \le 2|E|.|V|$$

اثبات. *فرض کنید T یک زیردرخت فراگیر از G باشد. به‌ازای هر رأس v می‌توانیم با شروع از v کل درخت را طی کنیم و دوباره به v برگردیم به‌طوری که هر یال دقیقاً ۲ بار طی شده‌باشد. این یک ویژگی درخت‌هاست که به سادگی از طریق استقرا قابل اثبات است یا به عنوان مثال می‌توان از الگوریتم جستجوی عمق‌اول [2] استفاده کرد. می‌توانیم رأس‌ها را بر اساس ترتیبی که پیموده می‌شوند شماره‌گذاری کنیم، $v_0 = v, v_1, ..., v_{2n-2} = v$. زمان پوشش برای v کمتر یا مساوی متوسط زمانی است که قدم‌زن تصادفی رأس‌ها را با این ترتیب می‌پیماید، یعنی ابتدا قدم‌زن با تعدادی گام از v_0 به v_1 می‌رسد، سپس با تعدادی گام از v_1 به v_2 و الی آخر. اثبات دقیق این موضوع به کمک بررسی درخت حالت‌های قدم‌زن تصادفی ممکن است. بنابراین*

$$\begin{aligned} C_v &\le \sum_{k=0}^{2n-3} H_{v_k, v_{k+1}} \\ &= \sum_{(x,y)\in E_T} H_{x,y} + H_{y,x} \\ &= \sum_{(x,y)\in E_T} 2|E|R(x,y) \\ &\le \sum_{(x,y)\in E_T} 2|E| \cdot 1 \\ &\le 2|E| \cdot |V| \end{aligned}$$

در رابطه‌ی آخر از این شهود فیزیکی استفاده کردیم که مقاومت معادل بین دو راس که بین آن‌ها یال هست، از مقدار مقاومت ۱ اهمی یال بین آن‌ها بیشتر نیست. این نکته به طور ریاضی نیز قابل اثبات است زیرا حل معادلات کرشهف با مینیمم‌سازی توان اتلافی الکتریکی معادل است. در نهایت با گرفتن ماکسیمم از دو طرف اثبات کامل می‌شود.

مراجع

[1] J. Laurie Snell Peter G. Doyle. Random walks and electric networks. 2006.

[2] Yuval Peres Russell Lyons. Probability on Trees and Networks. 2010.

[1] Cover Time

[2] Depth-First search

سه آزمون اول بودن با پیچیدگی زمانی چندجمله‌ای

سهیل معماریان

۱ مقدمه

تشخیص اول یا مرکب بودن یک عدد داده شده از بنیادی‌ترین مسائل نظریه اعداد الگوریتمی است که پژوهش در این رابطه تاکنون ادامه دارد. AKS اولین الگوریتم قطعی با پیچیدگی زمانی چندجمله‌ای برای این مقصود بود که در سال ۲۰۰۲ طراحی شد. این الگوریتم به عنوان یکی از مهم‌ترین دستاوردهای ۲۰ سال اخیر در مواجهه با این سوال محسوب می‌شود که علاوه بر نتایج درخشان آن، زیبایی و ظرافت اثبات‌هایی که در آن به کار گرفته شده نیز قابل توجه است.

در این مقاله که بر اساس [۲] نوشته شده است، با رهیافت نظری به بررسی آزمون AKS و دو آزمون اول بودن تصادفی (سولوی‌ـ استرسن و میلرـ رابین) پرداخته می‌شود به گونه‌ای که جز آشنایی با جبر و نظریه اعداد مقدماتی پیشنیاز دیگری فرض نشده است.

۲ قضیه‌ی کوچک فرما

در این بخش قضیه‌ی کوچک فرما و صورت‌های معادل آن را بیان می‌کنیم که در آزمون‌های اول بودن به‌کار می‌روند. این قضیه اولین بار توسط پیر دو فرما[1] در قرن هفدهم میلادی مطرح شد.

قضیه ۱. (*قضیه‌ی کوچک فرما*) *برای هر عدد اول p و هر عدد طبیعی a که $0 < a < p$ داریم:*

$$a^{p-1} \overset{p}{\equiv} 1.$$

اثبات ۱. چون $\mathbb{Z}_p^*$ یک گروه متناهی است، برای هر $a \in \mathbb{Z}_p^*$ داریم $\mathbb{Z}_p^* = a\mathbb{Z}_p^*$ ؛ در نتیجه حاصل‌ضرب اعضای آن‌ها نیز برابر است:

$$\prod_{b\in\mathbb{Z}_p^*} b \overset{p}{\equiv} \prod_{b\in\mathbb{Z}_p^*} ab \overset{p}{\equiv} a^{p-1} \prod_{b\in\mathbb{Z}_p^*} b \Rightarrow a^{p-1} \overset{p}{\equiv} 1.$$

لم ۱. *برای عدد صحیح a و عدد اول p اگر $a^2 \overset{p}{\equiv} 1$ آنگاه $a \overset{p}{\equiv} \pm 1$.*

اثبات ۲.

$$a^2 \overset{p}{\equiv} 1 \Rightarrow p|a^2 - 1 \Rightarrow p|(a-1)(a+1)$$

چون p اول است پس $p|(a-1)$ یا $p|(a+1)$.

قضیه ۲. *برای هر عدد اول فرد p و هر عدد طبیعی a که $0 < a < p$ داریم:*

$$a^{\frac{p-1}{2}} \overset{p}{\equiv} \pm 1$$

اثبات ۳. طبق قضیه ۱ داریم $a^{p-1} \overset{p}{\equiv} 1$. چون $p-1$ زوج است پس لم ۱ حکم را نتیجه می‌دهد.

قضیه ۳. *برای عدد اول فرد p که $p - 1 = 2^s t$ و t عدد طبیعی فرد باشد و عدد طبیعی a که $0 < a < p$، دنباله ی*

$$a^t (mod\ p), a^{2t} (mod\ p), a^{2^2 t} (mod\ p), ..., a^{2^s t} (mod\ p)$$

[1] Pierre de Fermat

یا تماماً ۱ است یا زوجِ ۱ـ و ۱ در جایی از دنباله ظاهر می‌شود و بعد آن تماماً ۱ است.

اثبات ۴. طبق قضیه ۱ آخرین عضو دنباله ۱ است. اگر تمام اعضای دنباله ۱ باشد که حکم ثابت می‌شود در غیر این صورت آخرین جایی را در نظر بگیرید که ۱ نیست و آنرا با a نشان می‌دهیم. چون هر عضو توان دوم عنصر قبلی است پس $a^2 \stackrel{p}{\equiv} 1$. بنابر لم ۱ باید $a \stackrel{p}{\equiv} -1$.

لم ۲. *عدد طبیعی n، عدد اول است اگر و فقط اگر برای هر $0 < i < n$ داشته باشیم $n|\binom{n}{i}$.*

اثبات ۵. می‌دانیم $\binom{n}{i} = \frac{n!}{(n-i)!i!}$. حال در صورت یک عامل n وجود دارد. اگر n عدد اول باشد آنگاه هیچ عاملی کوچکتر از خودش ندارد؛ لذا در مخرج هیچ عاملی از n وجود ندارد و یک طرف حکم ثابت می شود.

اگر n عدد مرکب باشد، عامل اولی مانند q دارد. k را عدد طبیعی تعریف می‌کنیم که $q^k|n$ و $q^{k+1} \nmid n$. $q^k \nmid \binom{n}{q}$ چون $\binom{n}{q} = \frac{n(n-1)...(n-q+1)}{q(q-1)...(1)}$ و تعداد عامل q در صورت برابر k و در مخرج برابر 1 است.

قضیه ۴. *عدد طبیعی n، عدد اول است اگر و فقط اگر برای هر a که $0 < a < n$ داشته باشیم:*

$$(x+a)^n \stackrel{n}{\equiv} x^n + a$$

(در اینجا منظور از همنهشت بودن دو چندجمله‌ای، همنهشت بودن ضرایب آن‌هاست.)

اثبات ۶. از اتحاد نیوتن می‌دانیم:

$$(x+a)^n - x^n - a = \sum_{i=1}^{n-1} \binom{n}{i} x^i a^{n-i} + a^n - a.$$

اگر n عدد اول باشد، طبق لم۲ ، $(x+a)^n \stackrel{n}{\equiv} x^n + a^n$. همچنین طبق قضیه ۱، $a^n \stackrel{n}{\equiv} a$. پس یک طرف حکم ثابت شد.

اگر n عدد اول نباشد، $n \nmid \binom{n}{i}$. اگر a را چنان انتخاب کنیم که $(n, a^{n-i}) = 1$ آنگاه ضریب جمله ی x^i در پیمانه ی n ناصفر می‌شود.

۳ آزمون سولوی‌ـاسترسن

آزمون سولوی‌ـاسترسن[2] اولین الگوریتم کارآمدی بود که برای آزمون اول بودن مطرح شد. ایده‌ی اصلی آن استفاده از قضیه‌ی ۲ و قضیه۷ است.

قضیه ۵. (*محک اویلر*[3]) *اگر p عدد اول فرد باشد آنگاه برای هر عدد طبیعی m داریم:*

$$(\frac{m}{p}) \stackrel{p}{\equiv} m^{\frac{p-1}{2}}.$$

اثبات ۷. اگر m بر p بخش‌پذیر باشد، هر دو طرف تساوی برابر صفر است. اگر $(p, m) = 1$ و g مولد گروه $\mathbb{Z}_p^*$ باشد، آنگاه عدد طبیعی i وجود دارد که $g^i \stackrel{p}{\equiv} m$.

اگر m مانده‌ی مربعی باشد، i زوج و طرف چپ تساوی برابر ۱ است. طبق قضیه‌ی ۱ داریم:

$$m^{\frac{p-1}{2}} \stackrel{p}{\equiv} g^{j\frac{p-1}{2}} \stackrel{p}{\equiv} 1 = (\frac{m}{p}).$$

اگر m نامانده‌ی مربعی باشد، i فرد و طرف چپ تساوی برابر ۱ـ است. طبق قضیه‌ی ۲ داریم $m^{\frac{p-1}{2}} \stackrel{p}{\equiv} \pm 1$. کافی است ثابت کنیم ۱ امکان‌پذیر نیست. چون g مولد است مرتبه‌ی آن $p-1$ است. اگر $1 \stackrel{p}{\equiv} m^{\frac{p-1}{2}} \stackrel{p}{\equiv} g^{j\frac{p-1}{2}}$، این نتیجه می‌دهد $p-1|j\frac{p-1}{2}$ که این با فرد بودن j در تناقض است.

تعریف ۱. (نماد ژاکوبی[4]) برای اعداد طبیعی m و n که تجزیه استاندارد n به صورت $n = \prod_{i=1}^{r} {p_i}^{\alpha_i}$ باشد، نماد ژاکوبی به صورت زیر تعریف می‌شود که منظور از $(\frac{m}{p_i})$ نماد لژاندار[5]است.

$$(\frac{m}{n}) = \prod_{i=1}^{r} (\frac{m}{p_i})^{\alpha_i}$$

ملاحظه ۱. از تعریف به سادگی بدست می‌آید $(\frac{a+n}{n}) = (\frac{a}{n})$.

[2] Solovay-Strassen
[3] Euler
[4] Jacobi
[5] Legendre

قضیه ۶. *(قانون تقابل مربعی برای نماد ژاکوبی) برای اعداد فرد* n, m *که* $(n,m)=1$*، داریم:*

$$(\frac{n}{m})(\frac{m}{n}) = (-1)^{\frac{(n-1)(m-1)}{4}}$$

اثبات آن را می‌توان در [۳] دید. توجه کنید که برای عدد مرکب n، لزومی ندارد که $(\frac{a}{n}) \overset{n}{\equiv} a^{\frac{n-1}{2}}$.

با استفاده از قضیه‌ی ۶ و ملاحظه‌ی ۱، با تقسیمات متوالی می‌توانیم الگوریتمی از مرتبه‌ی $O((log\ n)^2)$ برای محاسبه‌ی $(\frac{a}{n})$ بیان کنیم.

قضیه ۷. *n را عدد فردی در نظر بگیرید که توان کامل عدد اولی نباشد. عدد تصادفی a که $0<a<n$، یا $(a,n)>1$ یا با احتمال حداقل $\frac{1}{2}$ داریم:*

$$(\frac{a}{n}) \overset{n}{\not\equiv} a^{\frac{n-1}{2}}.$$

اثبات ۸. چون n مربع کامل نیست پس می‌توان نوشت $n=p^k m$ که p عدد اول و k عدد فرد باشد و نیز $(p,m)=1$. حال تعریف می کنیم:

$$A = \{0 < a < p^k | (a,p) = 1\}.$$

به وضوح $|A| = p^{k-1}(p-1)$ و دقیقاً $\frac{p^k(p-1)}{2}$ تااز اعضای A مانده‌ی مربعی و دقیقاً $\frac{p^k(p-1)}{2}$ تااز اعضای A نامانده‌ی مربعی هستند. اگر a_0،b_0 به ترتیب مانده‌ی مربعی و نامانده‌ی مربعی به پیمانه p در A باشند، عدد دلخواه c را که $0<c<m$ و $(c,m)=1$ را نیز در نظر بگیرید. با توجه به قضیه باقی‌مانده‌ی چینی، اعداد طبیعی یکتایی $0 < a,b < n$ وجود دارند که $a_0 \overset{p^k}{\equiv} a,\ b_0 \overset{p^k}{\equiv} b$ و $a \overset{m}{\equiv} b \overset{m}{\equiv} c$. حال داریم:

$$(\frac{a}{n}) = (\frac{a_0}{p})^k (\frac{c}{m}) = (\frac{c}{m})(*)$$

$$(\frac{b}{n}) = (\frac{b_0}{p})^k (\frac{c}{m}) = -(\frac{c}{m})(**)$$

اگر $(\frac{a}{n}) \overset{n}{\equiv} a^{\frac{n-1}{2}}$ و نیز $(\frac{b}{n}) \overset{n}{\equiv} b^{\frac{n-1}{2}}$ از $(*)$ و $(**)$ نتیجه می‌شود $a^{\frac{n-1}{2}} \overset{n}{\equiv} -b^{\frac{n-1}{2}}$. لذا خواهیم داشت:

$$c^{\frac{n-1}{2}} \overset{m}{\equiv} a^{\frac{n-1}{2}} \overset{m}{\equiv} -b^{\frac{n-1}{2}} \overset{m}{\equiv} -c^{\frac{n-1}{2}}$$

اما این امکان ندارد چون $(c,m)=1$. پس یا $(\frac{a}{n}) \overset{n}{\not\equiv} a^{\frac{n-1}{2}}$ یا $(\frac{b}{n}) \overset{n}{\not\equiv} b^{\frac{n-1}{2}}$. چون تعداد مانده‌های مربعی با تعداد نامانده‌های مربعی برابر بود پس برای عدد تصادفی a که $0<a<n$ و $(a,n)=1$ با احتمال حداقل $\frac{1}{2}$،

$$(\frac{a}{n}) \overset{n}{\not\equiv} a^{\frac{n-1}{2}}.$$

در هر سه آزمون اول بودنی که در این مقاله بررسی شده است ابتدا باید مطمئن باشیم که عدد داده شده توان کامل نیست. اما الگوریتم چک کردن این فرض دشوار نیست و به طور دقیق‌تر خواننده می‌تواند برای دیدن این الگوریتم به [۴] مراجعه کند. همچنین توجه کنید اگر صرفاً دنبال الگوریتم تصادفی برای آزمون اول بودن باشیم، اعدادی که به صورت توان کامل باشند برایمان مشکل‌ساز نخواهند بود زیرا ایناعداد در اعداد طبیعی چگالی صفر دارند و جمع معکوساتشان ۲ است:

$$\sum_{n=1}^{\infty}\sum_{k=2}^{\infty}\frac{1}{n^k} = 1 + \sum_{n=2}^{\infty}\frac{1}{n(n-1)} = 2$$

بنابر قضیه ۷ و قضیه ۲ می‌توان الگوریتم سولوی‌ـ استرسن برای آزمون اول بودن را مطرح کرد. این الگوریتم در صورتی که فقط یکبار اجرا شود به احتمال حداقل $\frac{1}{2}$ به جواب درست خواهد رسید. تعداد محاسبات مورد نیاز این الگوریتم، $O((log\ n)^3)$ است.

الگوریتم ۱. *(آزمون اول بودن سولوی‌ـ استرسن)*

input: n

output: prime or composite

if $\exists\, m,\ k \in \mathbb{N};\ n = m^k\ ,\ k>1$

return composite

End if

if $(\frac{a}{n}) = a^{\frac{n-1}{2}} (mod\ n)$

return prime

End if

return composite

۴ آزمون میلر ـ رابین

مایکل رابین[6] در [۶] با انجام تغیراتی در آزمون میلر[7] این الگوریتم را طراحی نمود. میلر با فرض درستی حدس ریمان تعمیم یافته[8] ثابت کرد برای عدد مرکب n که بیش از یک عامل اول داشته باشد، قضیه ۳ حداقل برای یک a که $0 < a < (log\ n)^2$ برقرار نیست. همچنین میلر ثابت کرد بدون پذیرفتن هیچ فرضی برای a تصادفی با احتمال بالایی قضیه‌ی ۳ برقرار نیست.

قضیه ۸. *اگر n عدد مرکبی باشد که بیش از یک عامل اول دارد و $n-1 = 2^s t$، با احتمال حداقل $\frac{1}{2}$، دنباله‌ی*

$$a^t(mod\ n), a^{2t}(mod\ n), a^{2^2 t}(mod\ n), ..., a^{2^s t}(mod\ n)$$

هیچ یک از این دو خاصیت را ندارد:

- *تماماً ۱ است.*
- *زوجِ ۱ـ و ۱ در جایی از دنباله ظاهر می‌شود و بعد آن تماماً ۱ است.*

اثبات ۹. p و q را دو عامل اول فرد n و k را بزرگترین توانی از p در نظر بگیرید که n را عاد می کند همچنین vوw اعداد طبیعی باشند که $p-1 = 2^v w$ و $(w,2) = 1$. ابتدا حالتی را بررسی می‌کنیم که حداقل یک ۱ـ در دنباله وجود داشته باشد. مجموعه A_u را برای $0 \le u < s$ به صورت زیر تعریف می‌کنیم:

$$A_u = \{a | (0 < a < n) \wedge (a^{2^u t} \overset{p^k}{\equiv} -1)\}.$$

پس برای هر $a \in A_u$ داریم $a^{2^u t} \overset{p^k}{\equiv} -1$. مجموعه $A_{p,u}$ را به صورت زیر تعریف می‌کنیم:

$$A_{p,u} = \{a\ (mod\ p^k) | a \in A_u\}.$$

چون تعداد اعضای گروه حاصل‌ضرب به پیمانه p^k، برابر $p^{k-1}(p-1)$ است برای هر $a \in A_{p,u}$ داریم $a^{p^{k-1}(p-1)} \overset{p^k}{\equiv} 1$. بنابراین $a^{(p^{k-1}(p-1), 2^{u+1}t)} \overset{p^k}{\equiv} 1$. p ، t را عاد نمی‌کند چون در غیر این صورت باید $n-1$ را عاد کند که امکان ندارد. لذا داریم $a^{(p-1, 2^{u+1}t)} \overset{p^k}{\equiv} 1$. چون t فرد است و $p-1 = 2^v w$ پس $a^{2^{min\{u+1,v\}}(w,t)} \overset{p^k}{\equiv} 1$. اگر $u \le v$ آنگاه خواهیم داشت $a^{2^u t} \overset{p^k}{\equiv} 1$ که امکان ندارد. پس $v > u$ که دو نتیجه در بردارد: اولاً $a^{2^u (w,t)} \overset{p^k}{\equiv} -1$؛ چراکه باید $a^{2^{u+1}(w,t)} \overset{p^k}{\equiv} 1$ (به لم ۱ توجه کنید.) و می‌دانستیم $a^{2^u t} \overset{p^k}{\equiv} -1$.

ثانیاً $2^{2u-v}(w,t) \le p-1$؛ چراکه $2u \le 2v$ و $(w,t) \le w$ نتیجه می‌دهد $2^{2u-v}(w,t) \le 2^v w = p-1$.

از سوی دیگر برای $l | p-1$ معادله $x^l \overset{p^k}{\equiv} -1$ حداکثر l جواب دارد لذا با توجه به دو نتیجه بیان شده داریم:

$$|A_{p,u}| \le 2^u (w,t) \le \frac{p-1}{2^{u-v}}.$$

با تکرار همین استدال برای $q-1 = 2^{v'} w'$ که $(w',2) = 1$ می‌توانیم بدست‌آوریم $|A_{q,u}| \le \frac{q-1}{2^{u-v'}}$ که $u < v'$ و $A_{q,u}$ به طور مشابه تعریف می‌شود. فرض کنید n فقط دو عامل اول p, q را داشته باشد. تعریف می‌کنیم $v'' = min\{v, v'\}$. از قضیه‌ی باقیمانده چینی داریم:

$$|A_u| < |A_{p,u}||A_{q,u}| \le \frac{p-1}{2^{u-v}} \frac{q-1}{2^{u-v'}} \le \frac{pq-1}{2^{2u-2v''}} \le \frac{n-1}{4^{u-v''}}.$$

بوضوح در صورتی که تعداد عوامل اول u بیشتر باشد نیز همین استدلال کارا است لذا برای $u < v''$، $|A_u| \le \frac{n-1}{4^{u-v''}}$. بنابراین

$$\sum_{0 \le u < s} |A_u| \le \sum_{0 \le u < v''} \frac{n-1}{4^{u-v''}} = (\frac{1}{3} - \frac{1}{3 \times 4^{v''}}).(n-1).$$

با توجه به تعریف A_u، $\frac{\sum_{0 \le u < s} |A_u|}{n-1}$ برابر است با احتمال اینکه در این دنباله در جایی ۱ـ بیاید و بعد آن ۱ باشد. لذا احتمال وقوع این احتمال حداکثر $\frac{1}{3} - \frac{1}{3 \times 4^{v''}}$ است. برای حالت تماماً ۱ نیز استدلال مشابه نتیجه می‌دهد تعداد a هایی که در دنباله ظاهر می‌شوند حداکثر برابر $\frac{1}{4^{v''}}(n-1)$ است. پس احتمال اینکه دنباله تولید شده حداقل یکی از دو خاصیت را داشته باشد کمتر $\frac{1}{3} - \frac{1}{3 \times 4^{v''}} + \frac{1}{4^{v''}}$ است که چون v'' حداقل ۱ است (عوامل اول فرد در نظر گرفته شده بودند.) این احتمال کمتر از $\frac{1}{2}$ است.

[6] Micheal Rabin
[7] Miller
[8] Extended Riemann hypothesis

آزمون اول بودن میلر–رابین با توجه به قضیه ۳ و قضیه ۸ با هر بار اجرا شدن با احتمال حداقل $\frac{1}{2}$ اول بودن یک عدد داده شده را درست تشخیص می‌دهد. این آزمون یک الگوریتم تصادفی با $O((log\ n)^2)$ محاسبه است.

الگوریتم ۲. *(آزمون اول بودن میلر– رابین)*

input: n

output: prime or composite

if $\exists\, m,\ k \in \mathbb{N};\ n = m^k\ ,\ k > 1$

return composite

End if

$a \leftarrow \{1, 2, ..., n\}$

$t \leftarrow n$

$k \leftarrow 0$

while $(t, 2) > 1$

$k \leftarrow k + 1$

$t \leftarrow \frac{t}{2}$

End while

if $(a, n) > 1$

return composite

End if

if $a^t = 1(mod\ n)$

return prime

End if

for $i = 0 : k$

$A[i] \leftarrow a^{2^i t}(mod\ n)$

if $i > 0, A[i] = 1, A[i-1] \neq -1$

return composite

End if

End for

return prime

۵ آزمون AKS

در سال ۲۰۰۴ کایال[9]و ساکسنا [10]این الگوریتم را در پروژه‌ی کارشناسی خود تحت نظر آگراول[11]طراحی کردند. برخلاف دو آزمون سولوی–استرسن و میلر–رابین، الگوریتم AKS قطعی است.

با استفاده از قضیه ۴ می‌توان یک آزمون به این صورت طراحی کرد که برای تشخیص اینکه n عدد اول است یا نه، به ازای همه‌ی $0 < a < n$ ها قضیه ۴ را امتحان کند. مشکلی که این آزمون دارد برای محاسبه‌ی ضرایب $(x + a)^n$ (یعنی $\binom{n}{i}$) به $\Omega(n)$ محاسبه نیاز دارد. برای حل کردن این مشکل قضیه‌ی ۴ را به‌این‌صورت بیان می‌کنیم:

گزاره ۱. *عدد طبیعی n اول است اگر و فقط اگر برای هر عدد طبیعی r و عدد طبیعی $a < n$ رابطه‌ی زیر برقرار باشد:*

$$(x + a)^n \equiv x^n + a\ (mod\ n, x^r - 1)$$

در قضیه‌ی ۱۰ ثابت خواهیم کرد که اگر گزاره ۱ برای r که به اندازه‌ی کافی کوچک انتخاب شده و تعداد کمی از aها برقرار باشد، آنگاه n باید عدد اول یا توانی از یک عدد طبیعی باشد. از طرفی چون r به اندازه‌ی کافی کوچک است، محاسبه ضرایب آسان می‌شود.

قضیه ۹. *اگر کوچکترین مضرب مشترک را با lcm نشان دهیم برای $n \geq 7$ داریم:*

$$lcm(1, 2, ..., n) \geq 2^n.$$

برای اثبات این قضیه می‌توانید به [۵] مراجعه کنید.

لم ۳. *برای هر عدد طبیعی n، عدد طبیعی r وجود دارد که مرتبه n در پیمانه r بیشتر از $log(n)^2$ است و $r \leq max\{3, (log\ n)^5 + 1\}$.*

(منظور از $log\ n$ ، لگاریتم n در مبنای دو است)

اثبات ۱۰. برای $n = 2$ و $r = 3$ حکم برقرار است. برای $n > 2$ حکم را بررسی می‌کنیم. چون $n > 2$ باید $(log\ n)^5 + 1 \geq 8$.

[9] Kayal
[10] Saxena
[11] Agrawal

$r_1, r_2, ..., r_m$ را تمام اعداد کمتر از n در نظر بگیرید که یا مقسوم علیه n هستند یا مرتبه n در آن پیمانه کمتر از $(log\ n)^2$ است. پس برای $1 \leq i \leq m$ داریم:

$$r_i | n. \prod_{i=1}^{i \leq (log\ n)^2} (n^i - 1).$$

همچنین

$$n. \prod_{i=1}^{i \leq (log\ n)^2} (n^i - 1) < n.n^{\sum_{i=1}^{i=(log\ n)^2}} < n^{(log\ n)^4} \leq 2^{(log\ n)^5}.$$

از طرفی طبق قضیه ۹ بزرگترین مضرب مشترک $(log\ n)^5$ عدد طبیعی ابتدایی بزرگتر از $2^{(log\ n)^5}$ است. پس عدد طبیعی کمتر از $(log\ n)^5$ مانند k وجود دارد که $k \notin \{r_1, r_2, ..., r_m\}$. اگر $(k, n) = 1$ باید مرتبه‌ی n به پیمانه k بیشتر از $(log\ n)^2$ باشد در غیر این صورت چون $k \nmid n$ پس $\frac{k}{(k,n)} \notin \{r_1, r_2, ..., r_m\}$ و در نتیجه مرتبه‌ی n به پیمانه $\frac{k}{(k,n)}$ بیشتر از $(log\ n)^2$ است.

قضیه ۱۰. *مرتبه‌ی عدد طبیعی n به پیمانه عدد طبیعی r بزرگتر از $(log\ n)^2$ است.(طبق لم ۳ ، چنین r ای وجود دارد.) اگر برای هر $0 \leq a \leq \sqrt{r} log\ n$ رابطه‌ی*

$$(x + a)^n \overset{n}{\equiv} x^n + a$$

برقرار باشد، n فقط یک عامل اول دارد.

اثبات ۱۱. p را عامل اولی از n در نظر بگیرید. دو مجموعه‌ی AوB را به صورت زیر تعریف می‌کنیم.

$$A = \{m | (x+a)^m = x^m + a (mod\ p, x^r - 1), 0 < a < \sqrt{r} log\ n\}$$

$$B = \{g(x) | g(x)^m = g(x^m) (mod\ p, x^r - 1), m \in A\}$$

با توجه به فرض، $p, n \in A$ و برای $0 \leq a \leq \sqrt{r} log\ n$، $x + a \in B$. از تعریف AوB روشن است که هر دو مجموعه نسبت به ضرب بسته هستند. در نتیجه هر دو مجموعه نامتناهی هستند. همچنین چون $p, n \in A$، برای اعداد طبیعی i و j، $(\frac{n}{p})^i p^j \in A$. دو مجموعه متناهی A_0 و B_0 را به صورت زیر تعریف می‌کنیم و با پیدا کردن کران برای $|A_0|$و$|B_0|$ حکم را ثابت می‌کنیم. چندجمله‌ای $h(x)$ دایره‌بر مرتبه r در $\mathbb{F}_p$ به عوامل تجزیه‌ناپذیری تجزیه می‌شود. $h(x)$ را یکی از این عوامل و $F = \frac{\mathbb{F}_p[x]}{(h(x))}$ در نظر بگیرید.

$$A_0 := \{m (mod\ r) | m \in A\}$$

$$B_0 := \{g(x) (mod\ p, h(x)) | g(x) \in B\}$$

ابتدا ثابت می‌کنیم $(log\ n)^2 < |A_0| < r$. $(n, r) = 1$ (چون در غیر این صورت n در پیمانه rمرتبه نخواهد داشت.) پس اعضای A_0 زیرمجموعه‌ای از $\mathbb{Z}_r^*$ هستند، لذا $|A_0| \leq \phi(r) < r$. همچنین چون مرتبه‌ی n در مرتبه‌ی r بزرگتر از $(log\ n)^2$ است و A_0 تمام توان‌های n را شامل است؛ درنتیجه $|A_0| > (log\ n)^2$.

حال ثابت می‌کنیم $2^{\sqrt{|A_0|} log\ n} \leq |B_0| \leq p^{r-1}$. چون اعضای B_0 چندجمله‌ایهای به پیمانه $h(x)$ هستند و درجه‌ی $h(x)$ کمتر از r است پس $|B_0| \leq p^{r-1}$.

برای اثبات کران پایین، ابتدا ثابت می‌کنیم دو چندجمله‌ای متمایز از درجه‌ی حداکثر $|A_0| - 1$ در B به دو چندجمله‌ای متمایز در B_0 نگاشته می‌شوند. فرض کنید چنین نباشد، پس $f(x), g(x) \in B$ وجود دارند که درجه‌ی آن‌ها کمتر از $|A_0|$ باشند و $f(x) = g(x) (mod\ p, h(x))$. این یعنی برای هر $m \in A_0$ داریم:

$$f(x^m) = f(x)^m = g(x)^m = g(x^m) (mod\ p, h(x)).$$

در نتیجه برای هر $m \in A_0$، x^m ریشه‌ی چندجمله‌ی $q(y) = f(y) - g(y)$ است. چون A_0 زیرمجموعه‌ای از $\mathbb{Z}_r^*$ بود، $(m, r) = 1$. پس هر چنین x^m ریشه‌ی rام واحد است. پس $q(y)$ حداقل $|A_0|$ ریشه در F دارد. اما از طرفی با توجه به درجه‌ی f و g، درجه‌ی q کمتر از $|A_0|$ است. این یعنی $f(x)$ و $g(x)$ دو چندجمله‌ای متحد روی B هستند.

تک‌جمله‌ای‌های $X, X+1, ..., X + \lfloor \sqrt{|r|}\ log\ n \rfloor$ در $F[X]$ متمایزند. حال هر جواب صحیح نامنفی از نامعادله‌ی $k_0 + k_1 + ... + k_{\lfloor \sqrt{|r|}\ log\ n \rfloor} \leq |A_0| - 1$ مانند $(k_0, k_1, ..., k_{\lfloor \sqrt{|r|}\ log\ n \rfloor})$ را متناظر کنید به چندجمله‌ی $(X)^{k_0} (X+1)^{k_1} ... (X + \lfloor \sqrt{|r|}\ log\ n \rfloor)^{k_{\lfloor \sqrt{|r|}\ log\ n \rfloor}}$. پس تعداد چندجمله‌ای‌های حداکثر از درجه $|A_0| - 1$ در $F[X]$ بیشتر از

$\binom{\lfloor\sqrt{r}\ log\ n\rfloor+|A_0|}{\lfloor\sqrt{r}\ log\ n\rfloor+1}$ است. باتوجه به کران پایینی که برای $|A_0|$ بدست آوردیم $|A_0| > \sqrt{|A_0|}log\ n$. لذا داریم:

$$\binom{\lfloor\sqrt{r}\ log\ n\rfloor+|A_0|}{\lfloor\sqrt{r}\ log\ n\rfloor+1} \geq \binom{\lfloor\sqrt{r}\ log\ n\rfloor+\sqrt{|A_0|}log\ n+1}{\lfloor\sqrt{r}\ log\ n\rfloor+1}$$
$$= \binom{\lfloor\sqrt{r}\ log\ n\rfloor+\sqrt{|A_0|}log\ n+1}{\sqrt{|A_0|}log\ n}$$

چون $r > |A_0|$ با جایگذاری در نامساوی بالا داریم:

$$\binom{\lfloor\sqrt{r}\ log\ n\rfloor+|A_0|}{\lfloor\sqrt{r}\ log\ n\rfloor+1} \geq \binom{2\lfloor\sqrt{|A_0|}\ log\ n\rfloor+1}{\lfloor\sqrt{|A_0|}\ log\ n\rfloor}$$
$$\geq 2^{\sqrt{|A_0|}\ log\ n} = n^{\sqrt{|A_0|}}.$$

در نتیجه $|B_0| > n^{\sqrt{|A_0|}}$.

برای $0 \leq i_1, i_2, j_1, j_2 \leq \sqrt{|A_0|}$ وجود دارد $(i_1, j_1) \neq (i_2, j_2)$ که $(\frac{n}{p})^{i_1}p^{j_1} \overset{r}{\equiv} (\frac{n}{p})^{i_2}p^{j_2}$ زیرا تعداد این زوج‌ها بیشتر از $|A_0|$ است و $(\frac{n}{p})^{i_1}p^{j_1}, (\frac{n}{p})^{i_2}p^{j_2} \in A$. همچنین چون $g(x) \in B_0$ داریم:

$$g(x)^{(\frac{n}{p})^{i_1}p^{j_1}} = g(x^{(\frac{n}{p})^{i_1}p^{j_1}}) = g(x^{(\frac{n}{p})^{i_2}p^{j_2}})$$
$$= g(x)^{n^{(\frac{n}{p})^{i_2}p^{j_2}}} (mod\ p, h(x)).$$

بنابراین $g(x)$ ریشه چندجمله‌ی $T(y) = y^{(\frac{n}{p})^{i_1}p^{j_1}} - y^{(\frac{n}{p})^{i_2}p^{j_2}}$ است. با توجه به کرانی که برای i_1, i_2, j_1, j_2 انتخاب شد، درجه‌ی چندجمله‌ای T حداکثر $n^{\sqrt{|A_0|}}$ است اما از طرفی تمام اعضای B_0 ریشه‌ی این چندجمله‌ای هستند. چون ثابت کرده‌ایم $|B_0| > n^{\sqrt{|A_0|}}$، این یعنی چندجمله T متحد با صفر است بنابراین باید $(\frac{n}{p})^{i_1}p^{j_1} = (\frac{n}{p})^{i_2}p^{j_2}$. در نتیجه داریم $n^{i_1-i_2} = p^{i_1+j_2-i_2-j_1}$ و این یعنی n توانی از یک عدد اول است.

آزمون AKS یک الگوریتم قطعی با کمتر از $O((log\ n)^5)$ محاسبه است که بنابر قضیه ۴، قضیه ۱۰ و لم ۳ آزمون توانایی تشخیص عدد اول از عدد مرکب را در زمان چندجمله‌ای دارد.

الگوریتم ۳. (*آزمون اول بودن* AKS)

input: n

output: prime or composite

if $\exists\, m,\ k \in \mathbb{N};\ n = m^k\ ,\ k > 1$

return composite

End if

$r \leftarrow$ the smallest r such that: $ord(n)\ \text{mod}\ r > (log(n))^2$

for $a = 0 : \lceil\sqrt{r}log\ n\rceil$

if $(n, a) \neq 1$ or $(x + a)^n \neq x^n + a(mod\ n,\ x^r - 1)$

return composite

End if

End for

return prime

مراجع

[1] M. Agrawal, N. Kayal, N. Saxena. PRIMES is in P. Annals of Mathematics, page 781–793, 2004.

[2] Manindra Agrawal. Primality tests based on fermat's little theorem.

[3] Tom M. Apostol. Introduction to Analytic Number Theory. deeplearningbook.org.

[4] Daniel J. Bernstein. Detecting perfect powers in essentially linear time. Mathematics of computation, 1998.

[5] Mohan Nair. On chebyshev-type inequalities for prime. Amer. Math. Monthly, page 126–129, 1982.

[6] Michael O. Rabin. Probabilistic algorithm for testing primality. Number Theory, page 128.

[7] Volker Strassen Robert M. Solovay. A fast monte-carlo test for primality. SIAM Journal on Computing, page 84–86, 1977.

همبستگی ماکسیمال

یاشار طالبی‌راد و حسین نادری

چکیده

در این مقاله ابتدا به بررسی علت مهم بودن همبستگی ماکسیمال [۱] و یکی از کاربردهای آن در تئوری اطلاعات اشاره می‌کنیم و به تعریف آن می‌پردازیم. سپس خواص ابتدایی آن را اثبات می‌کنیم و در انتها چند نامساوی پرکاربرد را بیان و اثبات می‌کنیم.

۱ مقدمه

یکی از مباحث مهم در تئوری اطلاعات، حفظ حریم داده[۲] است. مسأله‌هایی که در این مبحث مطرح می‌شوند را معمولا به این شکل شبیه‌سازی می‌کنند که دو نفر (مثل آلیس و باب) که از دو منبع اطلاعاتی مختلف (مثل X و Y) که توزیع توأم آن‌ها داده شده اطلاعات دریافت می کنند، آیا می‌توانند دو متغیر تصادفی A و B را که توزیع آن‌ها نیز داده شده است شبیه‌سازی کنند یا خیر؛ و اگر پاسخ بله است چگونه این کار را انجام دهند. معیارهای وابستگی[۳] مانند همبستگی ماکسیمال در پاسخ‌دادن به چنین سؤال‌هایی کمک می‌کنند.

۲ مدل‌سازی ریاضی

با توجه به این که شبیه‌سازی صد درصدی معمولا در کاربردهای واقعی غیرممکن است، در اکثر موارد کافیست فقط به مقدار لازم به توزیع A و B نزدیک شویم. یعنی آلیس و باب با انجام عملیات موضعی روی داده‌ها، متغیرهای تصادفی جدیدی می‌سازند که به A و B نزدیک باشند. عملیات موضعی روی متغیر تصادفی‌ای مثل X به معنی اعمال تابعی مانند f روی X و به‌دست آوردن $X' = f(X)$ است. حال فرض کنید آلیس و باب به‌جای این که فقط یک نمونه از X و Y داشته باشند، بتوانند به تعداد دلخواهی از آن نمونه تولید کنند؛ یعنی X^n و Y^n را داشته باشند. یک مدل پرکاربرد برای این مسأله، شبیه‌سازی غیرتعاملی[۴] نام دارد. شبیه‌سازی غیرتعاملی وقتی اتفاق می افتد که به ازای هر $\epsilon > 0$، توابع قطعی f, g وجود داشته باشند که $d_{\mathrm{TV}}\Big((f(X^n), g(Y^n)), (A, B)\Big) < \epsilon$، که در تعریف بالا d_{TV} برابر است با

$$d_{\mathrm{TV}}(A, B) = \sup_{T \in \mathcal{F}} \{|\mathcal{P}_A(T) - \mathcal{P}_B(T)|\}$$

قضیه ۱. *شبیه‌سازی غیرتعاملی امکان‌پذیر است اگر و فقط اگر (A, B) در بستار[۵] مجموعه‌ی زیر باشد:*

$$\Big\{(A', B') \mid A' \to X^n \to Y^n \to B' \text{ زنجیر مارکوف باشد}\Big\}$$

اما بررسی شرط فوق نیز آسان نیست چون نمی‌توان اعضای

[1] Maximal Correlation
[2] Data Privacy
[3] Measures of Dependence
[4] Non-interactive Simulation
[5] Closure

مجموعه‌ی بالا را به سادگی مشخص کرد. اما به‌جای مشخص‌کردن اعضا، می‌توان کران‌هایی پیدا کرد که اعضای آن مجموعه باید در آن کران‌ها صدق کنند. به این ترتیب اگر A و B و X و Y داده‌شده ای در آن کران‌ها صدق نکرد، می‌توان مطمئن بود که (A, B) در آن مجموعه نیست و بنابراین شبیه‌سازی غیرتعاملی امکان‌پذیر نیست. معیارهای وابستگی در یافتن این نوع کران‌ها کمک می‌کنند.

۳ معیارهای وابستگی

تاکنون معیارهای متعددی برای اندازه گیری میزان وابستگی دو متغیر تصادفی تعریف شده است. برای مثال یکی از پرکاربردترین آن‌ها، ضریب همبستگی پیرسون [۶]

$$\rho(X;Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

است. هر معیار وابستگی خوب است چند ویژگی زیر را داشته باشد:

۱. تحت عملیات موضعی عوض نشود.

۲. میزان وابستگی ۰ است اگر و فقط اگر دو متغیر مستقل باشند.

۳. میزان وابستگی ۱ است اگر و فقط اگر رابطه‌ی قطعی‌ای بین دو متغیر تصادفی وجود داشته باشد.

بنابراین خاصیت ۱ بیان می‌کند که وابستگی دو متغیر تصادفی در یک زنجیر مارکوف، یا گذر از یک کانال، یا انجام عملیات موضعی زیاد نشود. به طور دقیق‌تر برای هر معیار وابستگی C و هر زنجیر مارکوف $X \longrightarrow Y \longrightarrow Z$ باید داشته باشیم:

$$C(X;Z) \leq C(X;Y)$$

این نامساوی به نامساوی پردازش اطلاعات[۷] معروف است. با استفاده از این نامساوی، می‌توان به کران‌هایی در قضیه ۱ دست یافت. اما هر معیار وابستگی این ویژگی را ندارد. برای مثال ضریب همبستگی پیرسون ویژگی های ۱ و ۲ را ندارد.

یکی دیگر از این معیار های وابستگی، اطلاعات متقابل[۸] است. با استفاده از نامساوی پردازش اطلاعات برای اطلاعات متقابل در قضیه ۱ می‌توان گفت

$$I(A,B) \leq I(X^n;Y^n) = nI(X;Y)$$

اما سمت راست نامساوی وقتی $n \longrightarrow \infty$ برابر صفر یا بینهایت می‌شود. پس این معیار کمک چندانی نمی‌کند.

در [۳] بیان شده که یک خاصیت خوب دیگر می‌تواند خاصیت تانسوری[۹] باشد. گفته می‌شود معیار وابستگی C خاصیت تانسوری دارد وقتی $C(X^n;Y^n) = C(X;Y)$. به این ترتیب اطلاعات متقابل به علت جمعی[۱۰] بودن، تانسوری نیست. یک معیار مشابه که این مشکل را حل می‌کند، همبستگی ماکسیمال نام دارد که اثباتی از تانسوری بودن آن در [۵] ارائه شده است. تازگی ثابت شده است که مفهومی به نام نوار ابرانقباض پذیری[۱۱] کران‌های بهتری از همبستگی ماکسیمال نتیجه می‌دهد. به تعدادی از این کران‌ها در [۵] اشاره شده است.

۴ تعاریف اولیه

تعریف ۱. *همبستگی ماکسیمال بین دو متغیر تصادفی X, Y عبارت است از بیشینه‌ی ضریب همبستگی پیرسون بین $f(X)$ و $g(Y)$ روی همه توابع f و g.*

[6] Pearson Correlation Coefficient
[7] Data-Processing Inequality
[8] Mutual Information
[9] Tensorization
[10] Additive
[11] Hypercontractivity Band

به عبارت دیگر

$$\rho_m(X;Y) = \max_{f,g} \big|\rho(f(X);g(Y))\big| = \max_{f,g} \frac{\big|\mathrm{Cov}(f(X),g(Y))\big|}{\sigma_{f(X)}\sigma_{g(Y)}}$$

که در آن f و g توابع ثابت نیستند. وقتی یکی از X یا Y به طور قریب‌به‌یقین [12] ثابت باشد، همبستگی ماکسیمال ۰ تعریف می‌شود. همچنین Cov به معنی کوواریانس و σ_X همان انحراف معیار X است که به‌ترتیب عبارتند از

$$\mathrm{Cov}(X,Y) = \mathrm{E}\Big[(X-\mathrm{E}[X])\overline{(Y-\mathrm{E}[Y])}\Big] = \mathrm{E}[X\overline{Y}] - \mathrm{E}[X]\mathrm{E}[\overline{Y}]$$

$$\sigma_X = \sqrt{\mathrm{Var}(X)} = \sqrt{\mathrm{Cov}(X,X)} = \sqrt{\mathrm{E}\big[|X|^2\big] - \big|\mathrm{E}[X]\big|^2}$$

دقت کنید تعریف‌های بالا تعمیم حالت متغیرتصادفی حقیقی به حالت مختلط هستند.

۵ خواص

گزاره ۱. *برای هر دو متغیر تصادفی X,Y داریم*

$$0 \le \rho_m(X;Y) \le 1$$

برهان. طبق نامساوی کوشی‌ـشوارتز [13] داریم

$$|\mathrm{Cov}(Z,W)|^2 \le \mathrm{Var}(Z)\mathrm{Var}(W)$$

$$\Rightarrow |\mathrm{Cov}(Z,W)| \le \sqrt{\mathrm{Var}(Z)\mathrm{Var}(W)}$$

$$\Rightarrow \frac{|\mathrm{Cov}(Z,W)|}{\sqrt{\mathrm{Var}(Z)}\sqrt{\mathrm{Var}(W)}} \le 1 \Rightarrow |\rho(Z;W)| \le 1$$

با قرار دادن $Z = f(X), W = g(Y)$ طرف راست نامساوی ثابت می‌شود. طرف چپ بنابر تعریف بدیهی است. ■

برای بررسی چند خاصیت ابتدایی درباره‌ی استقلال، نیاز به تعریف تابع مشخصه[14] داریم. تابع مشخصه‌ی متغیر تصادفی X عبارت است از

$$\varphi_X(s) = \mathrm{E}[e^{isX}]$$

همچنین برای بردار تصادفی (X,Y) به این صورت تعریف می‌شود

$$\varphi_{X,Y}(s,t) = \mathrm{E}[e^{isX} \times e^{itY}]$$

تابع مشخصه مانند تابع توزیع احتمال، رفتار متغیر تصادفی را به طور یکتا مشخص می‌کند. به این معنی که X_1 و X_2 هم‌توزیع‌اند اگر و فقط اگر $\varphi_{X_1} = \varphi_{X_2}$.

لم ۱. *متغیر های تصادفی X و Y مستقل‌اند اگر و فقط اگر به‌ازای هر s و t داشته باشیم $\varphi_{X,Y}(s,t) = \varphi_X(s)\varphi_Y(t)$.*

برهان. ابتدا فرض کنید X و Y مستقل باشند. چون هر تابع‌هایی از متغیرهای مستقل خود نیز مستقل اند، برای هر t و s دو متغیر e^{itY} و e^{isX} نیز مستقل‌اند. در نتیجه

$$\mathrm{E}[e^{isX}e^{itY}] = \mathrm{E}[e^{isX}]\mathrm{E}[e^{itY}]$$

برای طرف دیگر ابتدا توجه کنید ضرب توزیع متغیرهای تصادفی $X \sim \mathcal{P}_X$ و $Y \sim \mathcal{P}_Y$ یعنی $P_X \times P_Y$ تابع مشخصه‌ای برابر $E[e^{isX}]E[e^{itY}]$ دارد که طبق فرض همان تابع مشخصه بردار تصادفی (X,Y) است. چون تبدیل فوریه [15] وارون‌پذیر است لذا تابع مشخصه یکتاست و $\mathcal{P}_X \times \mathcal{P}_Y = \mathcal{P}_{(X,Y)}$ که یعنی X و Y مستقل‌اند و حکم ثابت می‌شود. ■

قضیه ۲. *متغیر های تصادفی حقیقی X و Y مستقل‌اند اگر و فقط اگر $\rho_m(X;Y) = 0$.*

برهان. ابتدا فرض کنید X و Y مستقل باشند. چون هر تابع‌هایی از متغیرهای مستقل، خود نیز مستقل اند، برای هر f و g داریم $\mathrm{E}[f(X)\overline{g(Y)}] = \mathrm{E}[f(X)]\mathrm{E}[\overline{g(Y)}]$ و در نتیجه

$$\forall f,g: \qquad \rho(f(X);g(Y)) = \frac{\mathrm{Cov}(f(X),g(Y))}{\sigma_X\sigma_Y} = 0$$

[12] Almost Surely
[13] Cauchy-Schwarz Inequality
[14] Characteristic Function
[15] Fourier Transform

که یعنی همبستگی ماکسیمال برابر ۰ است.

برای طرف دیگر قضیه، اگر $\rho_m(X;Y)=0$ باشد، برای s و t ثابت قرار دهید $f(X)=e^{isX}$ و $g(Y)=e^{itY}$. توجه کنید که هرتابعی از e^{isX} تابعی از X نیز هست، و هر تابعی از e^{itY} تابعی از Y نیز هست پس بنابرتعریف

$$\rho_m(e^{isX};e^{itY}) \le \rho_m(X;Y)$$

ولی $\rho_m(e^{isX};e^{itY}) \ge 0$ و $\rho_m(X;Y)=0$ پس

$$\rho_m(e^{isX};e^{itY})=0 \Rightarrow \mathrm{Cov}(e^{isX},e^{itY})=0$$
$$\Rightarrow \mathrm{E}[e^{isX}\overline{e^{itY}}]=\mathrm{E}[e^{isX}]\mathrm{E}[\overline{e^{itY}}]$$

اما $\overline{e^{itY}}=e^{-itY}$ پس با تغییرمتغیر $t\leftarrow -t$ و طبق لم ۱، X و Y مستقل‌اند.

■

لم ۲. *برای هر دو تابع f و g می‌توان دو تابع f' و g' یافت به طوری که*

$$\mathrm{E}[f'(X)]=\mathrm{E}[g'(Y)]=0$$
$$\|f'(X)\|=\|g'(Y)\|=1$$
$$\rho\Big(f'(X);g'(Y)\Big)=\rho\Big(f(X);g(Y)\Big)$$

که در این جا $\|Z\|=\|Z\|_2=\sqrt{\mathrm{E}[|Z|^2]}$.

برهان. ابتدا فرض کنید $f'=f-\mathrm{E}[f(X)]$. در این صورت داریم

$$\rho(f'(X);g(Y))=\frac{\mathrm{Cov}(f(X)-\mathrm{E}[f(X)],g(Y))}{\sigma_{f(X)-\mathrm{E}[f(X)]}\sigma_{g(Y)}}$$

اما می‌دانیم واریانس و کوواریانس با اضافه شدن عددی ثابت به متغیرتصادفی تغییر نمی‌کنند. یعنی

$$\mathrm{Var}(Z+c)=\mathrm{Var}(Z), \qquad \mathrm{Cov}(Z+c,W)=\mathrm{Cov}(Z,W)$$

پس $\mathrm{E}[f'(X)]=0$ و

$$\rho(f'(X);g(Y))=\frac{\mathrm{Cov}(f(X),g(Y))}{\sigma_{f(X)}\sigma_{g(Y)}}=\rho(f(X);g(Y))$$

و به همین ترتیب می‌توان $\mathrm{E}[g'(Y)]$ را نیز بدون تغییر همبستگی برابر ۰ کرد. به این کار سنترینگ [16] می‌گویند. حال با فرض این‌که امیدریاضی $f(X)$ و $g(Y)$ صفر است قراردهید $f'=f/\|f(X)\|$. در این صورت

$$\rho(f'(X);g(Y))=\frac{\mathrm{Cov}(f(X)/\|f(X)\|,g(Y))}{\sigma_{f(X)/\|f(X)\|}\sigma_{g(Y)}}$$
$$=\rho(f(X);g(Y))$$

زیرا برای هر عدد حقیقی c داریم

$$\mathrm{Var}(cZ)=c^2\,\mathrm{Var}(Z), \qquad \mathrm{Cov}(cZ,W)=c\,\mathrm{Cov}(Z,W)$$

همچنین

$$\|f'(X)\|=\Big\|f(X)/\|f(X)\|\Big\|=\|f(X)\|/\|f(X)\|=1$$

و به همین ترتیب می‌توان $\|g'(Y)\|$ را نیز بدون تغییر همبستگی برابر ۱ کرد. به این کار اسکیلینگ [17] می گویند. به این ترتیب لم ثابت می‌شود.

■

با توجه به لم فوق، برای پیداکردن همبستگی ماکسیمال بین دو متغیر تصادفی، کافیست توجهمان را به توابع با امید ۰ و نرم ۱ و واریانس مثبت معطوف کنیم. برای سادگی به جای $f(X)$ از نماد f و به جای $g(Y)$ از g استفاده می‌کنیم. توجه کنید

$$\rho(f(X);g(Y))=\frac{\mathrm{E}[(f-\mathrm{E}[f])(g-\mathrm{E}[g])]}{\sqrt{(\mathrm{E}[f^2]-\mathrm{E}[f]^2)(\mathrm{E}[g^2]-\mathrm{E}[g]^2)}}$$
$$=\frac{\mathrm{E}[f(X)g(Y)]}{\sqrt{\|f\|^2\|g\|^2}}=\mathrm{E}[f(X)g(Y)]$$

پس کافیست $\mathrm{E}[f(X)g(Y)]$ را بیشینه کنیم. همچنین داریم

$$\mathrm{E}[f(X)g(Y)]=\mathrm{E}\Big[\mathrm{E}\big[f(X)g(Y)\big|Y\big]\Big]$$
$$=\mathrm{E}\Big[g(Y)\mathrm{E}[f(X)|Y]\Big]\le \|g(Y)\|\times\|\mathrm{E}[f(X)|Y]\|$$

که در گام آخر از نامساوی کوشی‌ـشوارتز استفاده شده‌است. حال دقت کنید $\|g(Y)\|=1$ و اگر f را داده‌شده فرض کنیم، برای بیشینه‌کردن همبستگی باید بین تمامی g ها آن تابعی را بیابیم که نامساوی بالا را به تساوی تبدیل کند. می‌دانیم حالت تساوی

[16] Centering
[17] Scaling

نامساوی کوشی–شوارتز وقتی اتفاق می‌افتد که دو متغیر تصادفی در راستای هم، یعنی ضریبی از هم باشند، یعنی

$$g(Y) = c\mathrm{E}[f(X)|Y]$$

اما باید $\|g\| = ۱$ و $\mathrm{E}[g] = ۰$ باشد. پس

$$g(Y) = \frac{\mathrm{E}[f(X)|Y]}{\|\mathrm{E}[f(X)|Y]\|}$$

و برای این f و g بهینه طبق نامساوی بالا داریم

$$\begin{aligned}\rho_m(X;Y) &= \max_{f:\mathrm{E}[f]=۰,\|f\|=۱} \|\mathrm{E}[f(X)|Y]\| \\ &= \max_{f:\mathrm{E}[f]=۰,\|f\|=۱} \sqrt{\mathrm{E}\Big[\mathrm{E}[f(X)|Y]^۲\Big]} \qquad \text{(I)}\end{aligned}$$

به همین ترتیب می‌توان برای $f(X)$ نیز رابطه مشابهی پیدا کرد

$$f(X) = \frac{\mathrm{E}[g^*(Y)|X]}{\|\mathrm{E}[g^*(Y)|X]\|}$$

که در این جا g^* جواب بهینه‌سازی زیر است

$$\rho_m(X;Y) = \max_{g:\mathrm{E}[g]=۰,\|g\|=۱} \|\mathrm{E}[g(Y)|X]\|$$

پس f و g بهینه که در همبستگی ماکسیمال صدق می‌کنند، باید در این روابط نیز صدق کنند، که به آن‌ها معادلات نقطه ثابت[18] می‌گویند. با استفاده از این روابط و الگوریتم امیدریاضی شرطی متناوب [19] می‌توان از یک حدس اولیه شروع کرد و به طور تکراری[20] به f و g بهینه میل کرد. منبع [۱] به این الگوریتم و کاربردهای آن به تفصیل پرداخته است.

برای حالت‌های خاصی از X و Y، روش‌های ساده‌ای برای محاسبه‌ی ρ_m ارائه شده است. برای مثال در [۴] ثابت شده اگر حداقل یکی از X و Y الفبای دوحرفی داشته باشد، داریم

$$\rho_m^۲(X;Y) = \left(\sum_{x,y} \frac{p(x,y)^۲}{p(x)p(y)}\right) - ۱$$

۶ کاربردها

همان‌طور که در ابتدای متن اشاره کردیم، یکی از کاربردهای همبستگی ماکسیمال در محاسبه کران‌هایی برای بررسی شبیه‌سازی غیرتعاملی است. چند کران پرکاربرد را در این قسمت بیان و اثبات می‌کنیم.

لم ۳. *فرض کنید $X \longrightarrow Y \longrightarrow Z$ یک زنجیر مارکوف باشد. چون Z و X به شرط Y مستقل‌اند، برای هر دو تابع f و g داریم*

$$\begin{aligned}\mathrm{E}[f(X)|Y]\mathrm{E}[g(Z)|Y] = \mathrm{E}[f(X)g(Z)|Y] \Rightarrow \\ \mathrm{E}[f(X)g(Z)] = \mathrm{E}[\mathrm{E}[f(X)g(Z)|Y]] \\ = \mathrm{E}[\mathrm{E}[f(X)|Y]\mathrm{E}[g(Z)|Y]]\end{aligned}$$

لم ۴. *اگر $X \longrightarrow Y \longrightarrow Z$ زنجیر مارکوف باشد، $\mathrm{E}[f(X)g(Z)] \le \rho_m(X;Y)\rho_m(Z;Y)$.*

برهان. طبق لم بالا داریم

$$\begin{aligned}\mathrm{E}[f(X)g(Z)] &= \mathrm{E}[\mathrm{E}[f(X)|Y]\mathrm{E}[g(Z)|Y]] \\ &\overset{(a)}{\le} \|\mathrm{E}[f(X)|Y]\| \times \|\mathrm{E}[g(Z)|Y]\| \overset{\text{طبق (I)}}{\le} \rho_m(X;Y)\rho_m(Z;Y)\end{aligned}$$

که در آن (a) از نامساوی کوشی–شوارتز نتیجه شده است. ■

یک نتیجه‌ی مهم این لم این است که

$$\begin{aligned}\rho_m(X;Z) &= \underset{\text{بهینه } f,g}{E}[f(X)g(Z)] \\ &\le \rho_m(X;Y)\rho_m(Z;Y) \le \rho_m(X;Y)\end{aligned}$$

که نامساوی آخر از گزاره ۱ نتیجه شده است. این نتیجه همان نامساوی پردازش اطلاعات برای همبستگی ماکسیمال است. از آن‌جایی که توزیع توأم X و Y را می‌توان به شکل یک کانال مخابراتی و یک توزیع ورودی برای آن نیز بیان کرد، گاهی به‌جای محاسبه‌ی همبستگی X و Y، همبستگی توزیع کانال و توزیع ورودی را بررسی می‌کنند. مثال‌هایی از این نوع بررسی در [۵] آمده

[18] Fixed-Point Equations

[19] Alternating Conditional Expectation algorithm

[20] Iterative

است. یک مسأله‌ی پرکاربرد در این خصوص، تکرار یک کانال است. یعنی $X \longrightarrow Y \longrightarrow Z$ که در آن (X,Y) با (Z,Y) هم‌توزیع باشد. یک نامساوی مهم برای این حالت را که از لم بالا نتیجه می‌شود این‌جا آورده‌ایم.

نتیجه ۱. *اگر $X \longrightarrow Y \longrightarrow Z \longrightarrow X$ زنجیر مارکوف باشد و (X,Y) با (Z,Y) هم‌توزیع باشد، داریم*

$$\rho_m^2(X;Y) = \max_{f:\mathrm{E}[f]=0,\|f\|=1} \mathrm{E}[f(X)f(Z)]$$

برهان. چون (X,Y) با (Z,Y) هم‌توزیع است داریم

$$\mathrm{E}[f(X)|Y] = \mathrm{E}[f(Z)|Y]$$
$$\Rightarrow \mathrm{E}[f(X)|Y]^2 = \mathrm{E}[f(X)|Y]\mathrm{E}[f(Z)|Y]$$
$$\Rightarrow \mathrm{E}\Big[\mathrm{E}[f(X)|Y]^2\Big] = \mathrm{E}\Big[\mathrm{E}[f(X)|Y]\mathrm{E}[f(Z)|Y]\Big]$$
$$= \mathrm{E}\Big[\mathrm{E}[f(X)f(Z)|Y]\Big] = \mathrm{E}[f(X)f(Z)]$$

اما طبق (I) داریم

$$\rho_m^2(X;Y) = \max_{f:\mathrm{E}[f]=0,\|f\|=1} \mathrm{E}[\mathrm{E}[f(X)|Y]^2]$$
$$= \max_{f:\mathrm{E}[f]=0,\|f\|=1} \mathrm{E}[f(X)f(Z)]$$

و حکم ثابت می‌شود.

دقت کنید چون

$$\max_{f:\mathrm{E}[f]=0,\|f\|=1} \mathrm{E}[f(X)f(Z)]$$
$$\leq \max_{f,g:\mathrm{E}[f]=\mathrm{E}[g]=0,\|f\|=\|g\|=1} \mathrm{E}[f(X)g(Z)]$$
$$= \rho_m(X;Z)$$

می‌توان نتیجه گرفت $\rho_m^2(X;Y) \leq \rho_m(X;Z)$. همچنین طبق نتیجه‌ی لم قبل داریم

$$\rho_m(X;Z) \leq \rho_m(X;Y)\rho_m(Z;Y) = \rho_m^2(X;Y)$$

که نتیجه می‌دهد $\rho_m(X;Z) = \rho_m^2(X;Y)$.

نتیجه ۲.

$$\sup_{\substack{X \longrightarrow Y \longrightarrow Z \\ \rho_m(Y;Z)\neq 0}} \frac{\rho_m(X;Z)}{\rho_m(Z;Y)} = \rho_m(X;Y)$$

توجه کنید طبق نتیجه‌ی لم ۴

$$\frac{\rho_m(X;Z)}{\rho_m(Z;Y)} \leq \rho_m(X;Y)$$

اثباتی برای رخ‌دادن تساوی در نامساوی بالا در [۴] ارائه شده‌است. یکی از توزیع‌های پرکاربرد برای X و Y، توزیع نرمال دومتغیره است. در [۲] همبستگی ماکسیمال برای این توزیع

$$\rho_m(X;Y) = |\rho(X;Y)|$$

محاسبه شده‌است. همچنین قضیه‌ی زیر را داریم

$$\rho_m^2(X;Y) \leq 1 - 2^{-2I(X;Y)} \leq (2\ln 2)I(X;Y)$$

اثباتی از این قضیه در [۴] آمده است. نکته‌ی جالب راجع به این نامساوی این است که همبستگی ماکسیمال را به اطلاعات متقابل ربط می‌دهد.

۷ نتیجه گیری

همان‌طور که دیدیم، همبستگی ماکسیمال در تئوری اطلاعات و حفظ حریم داده کاربردهای بسیاری دارد و کمک می‌کند بفهمیم چه موقع یک شبیه‌سازی امکان‌پذیر نیست.

۸ تشکر

در پایان از استادان گران‌قدر دکتر سلمان ابوالفتح بیگی و دکتر پویا شریعت‌پناهی که بدون رهنمودهای ایشان نگارش این نوشته ناممکن بود کمال سپاس‌گزاری را داریم.

مراجع

[1] J. Friedman L. Beriman. Estimating optimal transformations for multiple regression and correlation. Journal of the American Statistical Association, 1985.

[2] H. O. Lancaster. Some properties of the bivariate normal distribution considered in the form of the contingency table. Biometrika, Volume 44, Issue 1-2, page 289–292, 1957.

[3] A. Renyi. On measures of dependence. Acta Mathematica Hungarica.

[4] T. Linder S. Asoodeh, F. Alajaji. On maximal correlation, mutual information and data privacy. 2015 IEEE 14th Canadian Workshop on Information Theory (CWIT).

[5] Y. Park Z. Yin. Hypercontractivity, maximal correlation and non-interactive simulation.

دو فرهنگ ریاضی

علیرضا توکلی، علی چراغی

در سخنرانی مشهور سال ۱۹۵۹، با عنوان "دو فرهنگ"، اسنو[1] در مورد مضر بودن کمبود ارتباطات بین علوم انسانی و علوم طبیعی بحث کرد و از کسانی که روی علوم انسانی کار می‌کنند به دلیل عدم فهم علوم طبیعی انتقاد کرد. یکی از به یادماندنی‌ترین نقل‌قول‌ها که عدم تقارنی را نشان می‌دهد، در یک فرم متعادل‌تر، هنوز بعد از ۴۰ سال وجود دارد:

بسیاری از اوقات، من در جمعی از مردم بوده‌ام که، با استانداردهای جامعه سنتی، تحصیل‌کرده محسوب می‌شوند که با ذوق بسیاری ناباوریشان به بی‌سوادی دانشمندان را ابراز می‌کردند. یک بار یا دو بار، من تحریک شدم و پرسیدم که چند نفر از آنها می‌توانند قانون دوم ترمودینامیک را توضیح دهند. واکنش سرد بود و همچنین منفی بود. با این‌که، من در حال پرسیدن چیزی معادل علمی "آیا کاری از شکسپیر را خوانده‌اید؟" بودم.

من می‌خواهم در مورد یک اتفاق جامعه‌شناسانه مشابه که در ریاضیات محض دیده می‌شودو امر کاملاً سالمی نیست، صحبت کنم.

"دو فرهنگی" که من می‌خواهم در مورد آن بحث کنم، برای همه‌ی ریاضیدانان حرفه‌ای آشنا است. به طور غیردقیق، منظور من جدایی بین ریاضیدانانی است که هدف اصلی خود را حل مسئله می‌دانند، و آنهایی که بیشتر به ساختن و فهم نظریه‌ها می‌پردازند. این تفاوت نگرش توسط بسیاری از مردم اظهار شده است، و من هیچ ادعایی برای پی بردن به آن ندارم. مانند بسیاری از طبقه‌بندی‌ها، این نیز یک فراساده‌سازی را دربردارد، ولی نه آن‌قدر که آن را عبث کند. اگر شما مطمئن نیستید که در کدام دسته هستید، این دو عبارت را در نظر بگیرید.

۱) هدف حل مسئله این است که ریاضی را بهتر بفهمیم.

۲) هدف بهتر فهمیدن ریاضی این است که بهتر قادر باشیم مسائل را حل کنیم.

بیشتر ریاضیدان‌ها می‌گویند که درستی در هر دو (۱) و (۲) وجود دارد. همه مسائل به یک اندازه جالب نیستند، و یک راه تمیز دادن سوالات جالب‌تر این است که شرح دهیم که آن‌ها فهم ریاضی ما را در کل افزایش می‌دهند. به همان شکل، اگر کسی سال‌های زیادی را برای فهم قسمت سختی از ریاضی بگذراند، ولی هیچ کاری با این فهم انجام ندهد، چه اهمیتی برایش دارد؟ با این حال، بسیاری از ریاضیدانان با دو عبارت قبل به یک اندازه موافق نیستند. مایکل اتیه[2] یکی از ریاضیدانانی است که با این دو عبارت به یک اندازه موافق نیست، همان‌طور که در مصاحبه‌ای در سال ۱۹۸۴ نشان داده است [۲].

مصاحبه‌کننده: چگونه یک مسئله را برای مطالعه انتخاب می‌کنید؟

اتیه: من فکر می‌کنم این سوال، پیش‌فرضی را دربردارد. من فکر

[1] C.P. Snow

[2] Sir Micheal Atiyah

نمی‌کنم این همان راهی باشد که من کار می‌کنم. بعضی از مردم ممکن است بگویند "من می‌خواهم این سوال را حل کنم" و سپس آن‌ها با خود بگویند"چگونه این سوال را حل کنم؟"، من این کار را نمی‌کنم. من در آب‌های ریاضیات حرکت می‌کنم، در مورد چیزها فکر می‌کنم، با کنجکاو بودن، علاقه‌مند بودن، حرف زدن با مردم، دامن زدن به ایده‌ها؛ حقایق پدیدار می‌شوند و من آنها را دنبال می‌کنم. یا این‌که من چیزی می‌بینم که با دانسته های قبلی ام ارتباط دارد، و تلاش می‌کنم که آنها را کنار هم قرار دهم و پیشرفتی پدید می‌آید. من هیچگاه عملاً هیچ ایده‌ای را به این شکل که من چه کار می‌خواهم بکنم یا این ایده به کجا دارد می‌رود، شروع نکرده‌ام. من به ریاضی علاقه‌مندم، صحبت می‌کنم، یاد می‌گیرم، بحث می‌کنم و سپس به‌سادگی سوالات جالب پدیدار می‌شوند. من هرگز با هدف مشخصی شروع نکرده‌ام، بجز هدف فهم ریاضی.

این مصاحبه اولین بار در متمتیکال اینتلیجنسر[3] ظاهر شد، ولی در قسمت عمومی کارهای جمع شده‌ی اتیه نیز دوباره چاپ شد. من مقاله‌ها و سخنرانی‌های عمومی او را به هر کسی که می‌خواهد ایده‌هایش در مورد اهمیت ریاضی طبقه‌بندی کند، پیشنهاد می‌دهم.

شخص دیگری که روی این دو عبارت وزن یکسانی قرار نمی‌داد، پاول اردوش[4] بود، که جهانی از تعداد زیادی سوال‌های شگفت انگیز، به همراه جواب‌هایی شگفت انگیز به بسیاری دیگر به ارث گذاشت، ولی به همان اندازه با پیشرفت نظریه کاری نداشت. این به معنی انکار این نیست که اردوش در حال تلاش برای فهم ریاضیات بوده است: بسیاری از مردمی که سوالات اردوش راحل کرده‌اند (افسوس که من یکی از آنها نیستم)، شهادت می‌دهند که هر چقدر در مورد مسئله بیشتر فکر می‌کردند، در مسیرهای بسیار پرثمرتری قرار می‌گرفته‌اند و به این نتیجه رسیده‌اند که این مسئله از یک کنجکاوی جذاب بیشتر بوده است. پس وقتی من در مورد این صحبت می‌کنم که ریاضیدانها به دو دسته‌ی نظریه‌ساز و مسئله‌حل‌کن دسته‌بندی می‌شوند، من در حال صحبت در مورد اولویت‌های آن‌ها هستم، تا این ادعای مسخره که ریاضیدان ها منحصراً به یکی از دسته‌های کار ریاضی اختصاص داده شده‌اند.

واضح است که ریاضیات، هر دو نوع ریاضیدان را نیاز دارد (همان‌طور که اتیه در آخر [۱] می‌گوید) همچنین به همان اندازه واضح است که شاخه‌های مختلف ریاضی استعدادهای مختلفی نیاز دارند. در بعضی، مثل نظریه جبری اعداد، یا خوشه‌ای از موضوع‌ها که الان به‌سادگی هندسه نامیده می‌شوند، به نظر می‌آید (حداقل به عنوان یک بیگانه ــ من هیچ سندیتی برای حرف‌هایی که می‌زنم، ندارم) که به دلایل زیادی مهم است که فرد مقدار قابل توجهی تجربه و دانش در مورد کارهای دیگر ریاضیدانها داشته باشد، زیرا پیشرفت معمولاً حاصل ترکیب هوشمندانه برد وسیعی از نتایج موجود است. علاوه بر این، اگر کسی سوالی را انتخاب کند، و در انزوا چند سال روی آن کار کند و در آخر آن را حل کند، یک خطر وجود دارد، که شاید دیگر اهمیت قبل را نداشته باشد،مگر این‌که سوال بسیار معروفی باشد.

در طیف دیگر، برای مثال، نظریه گراف، که شیء ابتدایی، یک گراف، می‌تواند بلافاصله درک شود، یک نفر با نشستن بروی مبل و تلاش برای فهم بیشتر نظریه گراف به جایی نمی‌رسد. همچنین نیازی نیست که فرد قبل از این‌که به سوال حمله کند، مقدار زیادی از ادبیات آن را بداند. البته مفید است که از مهم‌ترین تکنیک‌های آن آگاه باشد، ولی سوالات جالب معمولاً به این دلیل باز هستند که تکنیک‌های بدست آمده را نمی‌توان به‌سادگی اعمال کرد.

اجازه دهید به‌طور مختصر یک عدم تقارن مشابه آن که اسنو به آن اشاره کرد را ذکر کنم. موضوع‌هایی که در حال حاضربه نظر نظریه‌سازها بسیار جذاب است بسیار مدروزتر از موضوع‌هایی است که برای مسئله‌حل‌کن‌ها جذاب است. علاوه بر این، ریاضیدانان در قسمت‌های نظریه‌سازی به عنوان کسانی که در حال انجام هسته اصلی هستند، دیده می‌شوند، در حالی که موضوع‌هایی مانند ترکیبیات، حاشیه‌ای و نامربوط به اهداف ریاضی دیده می‌شوند. کسی می‌تواند یک جمع از ریاضیدانهای بسیار تحصیل‌کرده را تصور کند که در حال ابراز ناباوری‌شان به بی‌سوادی ترکیبیات‌دان‌ها

[3] Mathematical Intelligencer

[4] Paul Erdos

[5] Calabi-Yau

هستند، که بسیاری از آن‌ها نمی‌توانند چیز هوشمندانه‌ای در مورد گروه‌های کوانتومی، تقارن آینه‌ای، منیفلدهای کالابی–یاو[۵]، معادله‌ی یانگ–میلز [۶]، یا حتی کوهمولوژی بگویند. اگر یک ترکیبیات‌دان، این صحبت را قطع کند و بپرسد چند زیرمجموعه از $\{1, 2, \cdots, n\}$ را می‌توان پیدا کرد که تفاضل متقارن هر دو تا از آنها اندازه حداقل $\frac{n}{3}$ داشته باشد، واکنش همچنین ممکن است کمی سرد باشد (این مسئله ساده است اگر و تنها اگر کسی تکنیک مربوطه را بداند، که این است که به طور تصادفی مجموعه‌ها را انتخاب کنیم و نشان دهیم که شانس این که هر جفتی از آنها تفاضل متقارن با اندازه کمتر از $\frac{n}{3}$ داشته باشد به‌طور نمایی کوچک است، پس جواب e^{cn} است برای $c > 0$ای.)

هدف من اینجا این است که از موضوع‌های کمتر مدروز در برابر انتقادهایی که در برابر آنها انجام می‌شود دفاع کنم. من مقدار زیادی از توجهم را به ترکیبیات اختصاص می‌دهم، زیرا این قسمتی است که از همه قسمت‌ها بهتر می‌شناسم. با این حال، چیزی که می‌گویم به بقیه قسمت‌ها نیز اعمال می‌شود.

من معمولاً کلمه‌ی "ترکیبیات" را به کار می‌برم نه به شیوه‌ای مرسوم، بلکه به عنوان یک کلمه عمومی که به سوالاتی اشاره کنم که معقول است از اولین اصول به آنها حمله کرد (این در واقع یک زاویه دید است تا یک جدایی کامل). این مسائل نیازی ندارند که گسسته باشند یا به شمارش ربط زیادی داشته باشند.

چرا باید موضوعات مسئله‌حل‌کردنی کمتر از نظری‌ها در نظر گرفته شوند؟ برای این که جواب این سوال را بدهیم باید یک سوال بنیادی‌تر را در نظر بگیریم: چه چیزی باعث می شود که قسمتی از ریاضی جالب‌تر از قسمت دیگری باشد؟ یک بار دیگر، اتیه بسیار واضح و جالب در مورد این موضوع می‌نویسد (در حالی که به ریاضیدانان بزرگ گذشته ادای دین می‌کند). او این نکته را خاطرنشان می‌کند (برای مثال [۱] را ببینید) که آن‌قدر ریاضیات زیاد تولید می‌شود که ممکن نیست همه آن را به خاطر نگاه داشت. پس روند مجردسازی و تعمیم، به عنوان معنی دادن به مقدار عظیمی داد ه خام (که اثبات قضیه‌های تکی هستند) بسیار مهم هست به طوری که این قابلیت را می‌دهد که مطالب انتقال یابند.

نتایجی که می‌مانند آن‌هایی هستند که می‌توانند به شکلی منسجم سازمان‌دهی شوند و به طور مقرون به صرفه به نسل‌های آینده ریاضیدانان توضیح داده شوند. البته، بعضی از نتایج به یاد می‌مانند به دلیل آن که می‌توانند مسئله‌های بسیار معروفی را حل کنند، و لی حتی اگراین نتایج نیزدر یک چارچوب سازمان‌دهنده قرار نگیرند، بعید است که به طور مفصل توسط ریاضیدانان مطالعه شوند.

پس، مفید است زیادتر در مورد جذابیت ذاتی نتیجه‌ی ریاضی بحث نکنیم تا این که چگونه این نتیجه می‌تواند به طور مؤثر به بقیه ریاضیدانان حاضر و آینده ابلاغ شود. ترکیبیات در نظر بسیاری، شامل تعداد زیادی از سوالات و نتایج ایزوله است، پس از این نظر، اشکال دارد. هر نتیجه به طور کلی ممکن است ابتکار بسیاری نیاز داشته باشد، و افراد مبتکر وجود دارند، ولی نسل‌های آینده‌ی ترکیبیات‌دان‌ها زمان یا تمایل این را ندارند که بیشتر از کسر کوچکی از این‌ها را بخوانند و تحسین کنند.

اجازه دهید که تلاش کنم به این انتقاد جواب دهم. قطعاً نادر است که در ترکیبیات کسی بتواند یک حکم بسیار کلی پیدا کند که ناگهان مقدار زیادی از نتایج موجود در بحث مربوطه را حل کند. همچنین این درست است که بسیاری از نتایجی که توسط ترکیبیات‌دان‌ها اثبات می‌شوند تا حدی ایزوله هستند و کاملاً فراموش خواهند شد (ولی این ترکیبیات را از شاخه‌های دیگر جدا نمی‌سازد). با این وجود، این درست نیست که به هیچ وجه ،ساختاری برای این موضوع وجود ندارد. دلیلی که به نظر بسیاری از ریاضیدانها، ترکیبیات یک مجموعه متفرقه از سوالات و نتایج تکی است، این است که اصول سازمان‌دهنده کمتر مشخص هستند.

اگر روند مجردسازی و تعمیم که در ریاضیات بسیار مهم هست، به طور محدود قابل استفاده در ترکیبیات می باشد، پس چگونه ممکن است که این موضوع به نسل‌های آینده انتقال داده شود؟ یک راه تفکر به این سوال این است که بپرسیم چه چیزهایی الزامات ترکیبیات‌دان‌های آینده خواهند بود؟ طبق چیزی که گفتم، اولویت آنها احتمالاً حل مسائل خواهد بود، پس علاقه آنها به نتایج امروز

[6] Yang-Mills equation

بسیار مربوط می‌شود به این که، آیا با فهمیدن آن، می‌توانند قدرت حل مسئله خود را افزایش دهند یا نه. و این‌ها ما را سرراست به اصل مطلب می‌برند. ایده‌های مهم ترکیبیات معمولاً به طور یک قضیه مشخص ظاهر نمی‌شوند، بلکه معمولاً به عنوان یک اصل عمومی با کاربرد بسیار ظاهر می‌شوند.

یک مثال کمک می‌کند که این نکته مبرهن‌تر شود. یک فرم قضیه رمزی[7] عبارت زیر است:

قضیه. برای هر عدد صحیح مثبت k، یک عدد صحیح مثبت N وجود دارد که اگر یال‌های یک گراف کامل n رأسی قرمز یا آبی شوند، آنگاه حتماً k رأس وجود دارند که یال‌های وصل کننده آن‌ها همگی یک رنگ دارند.

کوچکترین N را با $R(k)$ می‌شناسند.

بحث زیر نشان می‌دهد که $R(k) \leq 2^{2k}$. فرض کنید G گرافی با 2^{2k} رأس باشد و برای راحتی فرض کنید که رئوس کاملاً مرتب شده هستند. فرض کنید x_1 رأس اول باشد، آنگاه طبق اصل لانه کبوتری یک مجموعه از رئوس، A_1، از اندازه حداقل 2^{2k-1} وجود دارد که هر یال از x_1 به A_1 یک رنگ را دارد. حال فرض کنیم x_2 کوچکترین رأس A_1 باشد. آنگاه طبق اصل لانه کبوتری $A_2 \subseteq A_1$ از اندازه حداقل 2^{2k-2} وجود دارد که هر یال از x_2 به A_2 یک رنگ دارد. با ادامه دادن این روند، یک دنباله از رئوس $x_1, x_2, \cdots, x_{2k}$ و یک دنباله $A_1 \subseteq \cdots \subseteq A_{2k-1} \subseteq A_{2k}$ از مجموعه‌ها پیدا می‌کنیم که $x_i \in A_{i-1}$ برای هر i و هر یال از x_i به A_i یک رنگ دارد. بدست می‌آید که رنگ یال متصل‌کننده x_i به x_j فقط به $\min\{i,j\}$ بستگی دارد. دوباره با استفاده از اصل لانه کبوتری، یک زیرمجموعه H از $\{x_1, x_2, \cdots, x_{2k}\}$ پیدا می‌کنیم که این رنگ همیشه یکسان باشد، پس تمام یال‌های وصل‌کننده رئوس H یک رنگ دارند.

یک تغییر کوچک در استدلال بالا کران بالا را به $\binom{2k}{k}$ بهبود می‌بخشد. با این وجود، من بیشتر به کران‌های *پایین* $R(k)$ علاقه‌مندم. یکی از نتایج مشهور اردوش این نتیجه است که $R(k) \geq 2^{k/2}$ و بدین شکل اثبات می‌شود. به جای تلاش برای ساختن یک رنگ‌آمیزی هوشمندانه، یکی به سادگی یال‌ها را به طور تصادفی در واضح‌ترین راه رنگ می‌کند که این است که هر یال با احتمال $\frac{1}{2}$ قرمز است و با احتمال $\frac{1}{2}$ آبی است و همه این انتخاب‌ها مستقل‌اند. فرض کنید تعداد رئوس N باشد، و $\{x_1, x_2, \cdots, x_k\}$ kتا از آن‌ها باشند. احتمال اینکه هر x_i به هر x_j با یک یال قرمز وصل شده باشد $2^{-\binom{k}{2}}$ است، و برابر این است که همگی با یال‌های آبی وصل شده باشند. پس مقدار مورد انتظار برای مجموعه‌های k رأسی که همگی آن‌ها با یال‌های همرنگ به هم وصل باشند $2^{1-\binom{k}{2}}\binom{n}{k}$ است. اگر این کمتر از ۱ باشد، باید ممکن باشد که هیچ چنین زیرمجموعه‌ی k رأسی برای آن موجود نباشد. یک محاسبات کوچک نشان می‌دهد که این کمتر از ۱ است اگر $N = 2^{k/2}$.

نتیجه‌ی اردوش [۵] مشهور است نه به این دلیل که کاربردهای بسیاری دارد، نه به دلیل اینکه سخت است، و نه به دلیل اینکه یک سوال باز طولانی‌مدت را حل کرد. شهرت آن به این دلیل است که راه‌های عظیم استدلال‌های احتمالاتی را به ترکیبیات باز کرد. اگر شما استدلال ساده‌ی اردوش را می‌فهمید (یا یکی از بسیار استدلال‌های مشابه) آنگاه در ذهن شما یک اصل کلی به همراه خط‌های زیر ایجاد می‌شود:

اگر کسی بخواهد اندازه‌ی یک ساختار را تحت تعدادی محدودیت ماکسیمم کند، و اگر محدودکننده‌ها به نظر مثال‌های اکسترمال را مجبور کنند که به شکلی یکنواخت پخش شوند، آنگاه انتخاب یک مثال به طور تصادفی به نظر می‌آید جواب خوبی دهد.

همین که شما از این اصل آگاه می‌شوید، قدرت ریاضی شما بلافاصله افزایش می‌یابد. سوالاتی مانند آن که قبلاً گفتم در مورد پیدا کردن تعداد زیادی از مجموعه‌ها با تفاضل متقارن‌های بزرگ ناگهان از غیرممکن به تقریباً بدیهی تبدیل می‌شوند.

البته، بیشتر از این کاربرد ساده در مورد ترکیبیات وجود دارد. برای مثال ممکن است کسی تصمیم بگیرد که از روش‌های احتمالاتی استفاده کند و بعد بفهمد که تخمین زدن احتمال‌های مربوطه کار آسانی نیست. با این وجود مقدار زیادی کار در این مورد انجام

[7] Ramsey

شده، و تکنیک‌های هوشمندانه بسیاری ساخته شده‌اند. بعضی از این‌ها دوباره در قانون‌های مفیدی محصور شده‌اند. یکی از آن‌ها، که به دوستانش به عنوان تمرکز اندازه[8] شناخته شده است، عبارت زیر است:

اگر یک تابع f به شکل قابل قبول پیوسته‌ای به تعداد زیادی متغیر کوچک وابسته باشد، آنگاه $f(x)$ تقریباً همیشه به مقدار مورد انتظارش نزدیک است.

اهمیت کامل تمرکز اندازه اول توسط ویتالی میلمن[9] در اثبات انقلابی‌اش [۱۱] از قضیه زیر از دورتسکی[10] [۴] پی برده شد.

قضیه. برای هر عدد صحیح مثبت k و $\epsilon > 0$ یک n هست که هر فضای نرمدار n بعدی X یک زیرفضای k بعدی با فاصله باناخ–میزر[11] حداکثر $1+\epsilon$ تا l_2^k دارد.

این معادل این است که بگوییم می‌توان بردارهای $x_1, x_2, \cdots, x_k \in X$ [مستقل خطی] پیدا کرد که

$$(\sum_{i=1}^{k} |a_i|^2)^{\frac{1}{2}} \leq ||\sum_{i=1}^{k} a_i x_i|| \leq (1+\epsilon)(\sum_{i=1}^{k} |a_i|^2)^{\frac{1}{2}}$$

برای هر دنباله $a_1, a_2, \cdots, a_k$ از اسکالرها. یک راه هندسی‌تر (و حتی غیر شهودی‌تر) برای دوباره فرمول‌بندی این قضیه این است که بگوییم هر جسم محدب متقارن n بعدی، یک مقطع مرکزی k بعدی دارد که یک بیضی‌گون k بعدی B را در بر دارد و در $(1+\epsilon)B$ قرار دارد و پس خود آن نیز تقریباً بیضی‌گون است.

رویکرد میلمن به این قضیه این است که یک زیرفضای k بعدی از X را بطور تصادفی انتخاب کنیم. قبل از انجام این کار، اول باید یک اندازه احتمال با معنی انتخاب کرد، که می‌تواند با استفاده از قضیه‌ای از فریتز جان[12] انجام شود. این بیان می‌کند که یک پایه $x_1, x_2, \cdots, x_n$ از X وجود دارد که به شکل ناامیدکننده‌ای دور از متعامد یکه نیست، به این معنی که:

$$(\sum_{i=1}^{n} |a_i|^2)^{\frac{1}{2}} \leq ||\sum_{i=1}^{n} a_i x_i|| \leq \sqrt{n}(\sum_{i=1}^{n} |a_i|^2)^{\frac{1}{2}}$$

برای هر دنباله $a_1, a_2, \cdots, a_k$ از اسکالرها. حال یکی اندازه عادی روی گراسمانیان‌های $G_{n,k}$ نسبت به این پایه را قرار می‌دهد (در واقع این یک فراساده‌سازی است، ولی جزئیات دقیق را کاری نداریم).

در حالت دو نتیجه‌ای که قبلاً به آن اشاره کردم، راحت بود که ببینیم، که روش‌های تصادفی با معنی بودند. بعد از این همه، ممکن است کسی فکر کند که یک مقطع نوعی از یک جسم محدب غیرمنظم، غیرمنظم خواهد بود. با این وجود، همین که یک نفر با ایده‌ی تمرکز اندازه آشنا باشد، به این پی می‌برد که نرم بردار $\sum_{i=1}^{n} a_i x_i$ به متغیرهای چندگانه a_i وابسته است. همچنین، چون ما فقط به نسبت این نرم به نرم l_2 علاقه‌مندیم، می‌توانیم فرض کنیم که هیچ کدام از a_iها تغییر نمی‌کنند. پس، با تمرکز اندازه، ممکن است انتظار داشته باشیم که نسبت نرم X به نرم l_2 خیلی به مقدار مورد انتظار خود در همه زمان‌ها نزدیک باشد. و الآن ایده‌ی مقاطع تقریباً بیضی‌گون به یک ضدشهود منجر شده است.

نتیجه اصلی برای این که حکم بالا را دقیق کنیم، این نتیجه از نامساوی برابر محیطی[13] لوی[14] روی کره است. تابع $f : S_n \to \mathbb{R}$ را تابعی با میانه M قرار دهید. فرض کنید $A \subseteq S_n$ مجموعه‌ی همه‌ی نقاط x باشد که $f(x) \leq M$. آنگاه احتمال اینکه یک نقطه‌ی تصادفی در S_n فاصله بیشتر از ϵ تا A داشته باشد، حداکثر $\sqrt{\frac{\pi}{2}} \exp(\frac{-\epsilon^2 n}{2})$ است. حال قرار دهید $f(a_1, a_2, \cdots, a_n) = ||\sum_{i=1}^{k} a_i x_i||$. چون f به شکلی قابل قبول پیوسته است، و پس چون اکثر نقاط y به نقاط $x \in A$ نزدیک‌اند، پس نتیجه می‌شود که $f(y)$ خیلی بزرگتر از M نیست. به طور مشابه، برای اکثر yها،

[8] Concentration of measure
[9] Vitali Milman
[10] Dvoretzky
[11] Banach-Mazur distance
[12] Fritz John
[13] Isoperimetric inequality
[14] Levy

$f(y)$ خیلی کمتر از M نیست.

قضیه دورتسکی، به خصوص با اثبات میلمن، یک سنگ بنا در نظریه موضعی (به معنی متناهی بعد) فضاهای باناخ است. در حالی که متأسفم برای ریاضیدانی که نمی‌تواند جذبه ذاتی را ببیند، این جذبه به خودی خود، اثر عظیم اثبات را، که بالاتر از نظریه فضاهای باناخ، به عنوان نتیجه‌ای از قرار دادن ایده‌ی تمرکز اندازه در ذهن بسیاری ریاضیدان داشته است، توضیح نمی‌دهد. تعداد بسیاری از مقالات این ایده را باز کرده‌اند یا تکنیک‌های جدیدی برای این‌که این اتفاق می‌افتد، نشان داده‌اند.

می‌توان اصول کلی بیشتری از این دست را نام برد. در اینجا چندتا را نوشته‌ام. اهمیت این اصل‌ها با هم برابر نیستند و همان‌گونه که گفتم دقیق نیستند ولی همه‌ی آنها به مقدار زیادی بررسی شده اند.

۱) به وضوح اگر پیشامدهای $E_1, E_2, \cdots, E_n$ مستقل از هم باشند و دارای احتمال ناصفر باشند، در این صورت با احتمال ناصفری همه‌ی این پیشامدها با هم اتفاق می‌افتند. در واقع این حکم می‌تواند در حالتی که این پیشامدها وابستگی کمی به هم داشته باشند نیز مورد استفاده قرار گیرد. [۶]،[۹]

۲) همه‌ی گراف‌ها از چند قسمت تصادفی‌گون تشکیل شده‌اند و ما می‌دانیم که این قسمت‌ها چگونه رفتار می‌کنند. [۱۴]

۳) اگر داریم تعداد جواب‌های یک معادله خطی را درون یک مجموعه می‌شماریم، کافیست و معمولاً ساده‌تر است که ضرایب فوریه تابع مشخصه این مجموعه را تخمین بزنیم.

۴) بسیاری از خواص گراف‌های تصادفی با هم هم‌ارز هستند و نتیجتاً می‌توان آنها را به عنوان تعاریف گراف‌های شبه تصادفی درنظر گرفت. [۷]،[۱۵]

۵) گاهی اوقات مجموعه دنباله‌های در نهایت صفر از صفر و یک‌ها مدل خوبی برای یک فضای باناخ جدایی‌پذیر است و یا حداقل به فرد اجازه می‌دهد که فرضیات جالب توجهی را تولید کند. نکته اصلی که در مورد این گونه اصول می‌خواهم بگویم این نیست که اینها بسیار پرکاربرد هستند؛ که البته نکته تعجب‌برانگیزی هم نیست؛ بلکه این است که اصول نقش جهت‌دهنده‌ای را در ترکیبیات دارند که قضایای عمیق و کلی در زمینه‌های نظری دیگر دارند. وقتی برای فهمیدن نتیجه‌ای تلاش می‌کنیم زمان زیادی را می‌توانیم صرفه‌جویی کنیم اگر بتوانیم نتیجه را به دو یا سه ایده کلی تحویل دهیم. بعد از انجام این کار ممکن است که نیازی به بررسی دقیق جزئیات نبینیم. آشنایی کلی با اصول کلی و چند مثال از کاربردهای آن باعث می‌شود که فرد بتواند در صورت نیاز، خود جزئیات را انجام دهد. برای مثال من برهان اردوش از کران پایین برای $R(k)$ را این‌گونه به خاطر سپرده‌ام: رنگ آمیزی گراف را به صورت تصادفی انجام دهید و محاسبات بدیهی را انجام دهید. یک مثال دیگر قضیه قابل توجه زیر از کاشین است. [۱۰]

قضیه. برای هر عدد صحیح n یک تجزیه عمود از $\mathbb{R}^{2n}$ (نسبت به ضرب داخلی معمولی) به دو زیر فضای n بعدی X و Y وجود دارد به طوری که برای هر $x \in X \cup Y$ نسبت نرم $x\ l_1$ به نرم $x\ l_2$ بین $c\sqrt{2n}$ و $\sqrt{2n}$ برای ثابت مطلق $c > 0$.

برهانی از زارک[15] [۱۳] از این قضیه به شکل زیر است (اگر با ایده های مبهم درستی آشنا باشید): یک استدلال ساده بر مبنای حجم نشان می‌دهد که تقریبا هیچ‌یک از گوی‌های یکه‌ی l_1^n در گوشه‌ها قرار ندارد (منظور از گوشه، قسمت‌هایی است که نسبت l_1ـنرم به l_2ـنرم کوچک است)، پس کار ساده‌ایست که نشان دهیم یک تجزیه رندم خاصیت موردنظر را دارد. مثال سوم قضیه زیر از رات[16] است. [۱۲]

قضیه. برای هر δ، عدد طبیعی N وجود دارد که هر زیرمجموعه از $\{1, 2, \cdots, N\}$ به اندازه حداقل δN (یعنی با چگالی حداقل δ) شامل یک تصاعد حسابی به طول 3 باشد.

برهان فشرده این حکم از قرار زیر است. به $A \subseteq \mathbb{Z}/N\mathbb{Z}$ نگاه کنید و ضرایب فوریه تابع مشخصه A را در نظر بگیرید. اگر این ضرایب کوچک باشند (به غیر از $\hat{A}(0)$) در این صورت A عملاً تصادفی خواهد بود، و در نتیجه شامل تعداد زیادی تصاعد حسابی به طول 3 خواهد بود. در غیر این صورت ضریب بزرگی وجود خواهد داشت که به ما اجازه می‌دهد زیرتصاعدی از $\{1, 2, \cdots, N\}$ را بیابیم که A در آن دارای چگالی بیشتری است.

[15] Szarek
[16] Roth

اجازه دهید مطالبی که تاکنون گفته‌ام را خلاصه کنم. تاکنون سعی کرده‌ام علیه این ایده که موضوع ترکیبیات ساختار کمی دارد و فقط از تعداد زیادی مسئله تشکیل شده است، استدلال کنم. در حالی که این ساختار نسبت به موضوعات دیگر کمتر جلوه می‌کند، این ساختارها به صورت جملاتی کلی و مبهم وجود دارند که اجازه می‌دهند که ما برهان‌ها را به صورت فشرده‌ای در ذهن خود نگاه داریم و در نتیجه راحت‌تر به خاطر بسپاریم و به دیگران منتقل کنیم.

با این حال انتقادات زیادی به ترکیبیات وارد است. یک مورد که من شنیده‌ام این است که این موضوع جهت‌دهی و اهداف کلی ندارد. یکی دیگر این است که این موضوع عمیق نیست و دیگر این که ترکیبیات ارتباطات جالبی به بقیه بخش‌های دیگر ریاضی (یعنی بخش‌های محوری ریاضیات) ندارد و دیگر این که خیلی از نتایج دارای کاربرد نیستند.

این انتقادات را می‌توان به شیوه‌ی مشابهی پاسخ داد. مثلاً این نکته را در نظر بگیرید که اهداف کلی در ترکیبیات وجود ندارد. دوباره، من از یک مصاحبه با اتیه نقل می‌کنم [۲]: من داشتم در مورد گرایش امروزه‌ی افراد به ایجاد یک شاخه‌ی کامل ریاضی به طور شخصی و به طور مجرد فکر می‌کردم. این‌ها فقط خشت بر می‌زنند و اگر از آنها بپرسیم که به خاطر چه این کار را انجام می دهند، اهمیت این موضوع در چیست، به چه مربوط است، متوجه خواهید شد که خودشان نمی‌دانند.

منظور اتیه به طور خاص ترکیبیات نبود ولی او به نکته‌ی مهمی اشاره می‌کند و این موضوع برای ترکیبیات‌دان‌ها حائز اهمیت است همان‌گونه که برای هر کس دیگری این گونه است که نشان دهد که کاری را که انجام می‌دهند بیش از زیاده‌کاری است. بعضی از بخش‌های ریاضیات هستند که تحت سلطه‌ی تعداد کمی مسئله که دارای اهمیت جهانی هستند، قرار گرفته‌اند. می‌توان خیلی از نتایج را این گونه توجیه کرد که آنها هر چند به طور جزئی در مورد فرضیه ریمان، حدس بیرچ سوینرتون‌ـ دایر، حدس هندسی‌سازی ترستن، حدس نوویکوف یا چنین چیزی روشنگری می‌کنند. شاخه‌های دیگر ریاضی جذابیت خود را از فراوانی پدیده‌های رمزآلود که نیازمند توضیح هستند، بدست می‌آورند. ممکن است که همرویدادهای عددی باشند که ارتباط عمیقی را بین حوزه هایی که در نگاه اول بی‌ارتباط به نظر می‌رسند پیشنهاد کنند، استدلال‌هایی که نتایج جالبی را با توسل به محاسبه اثبات می‌کنند و در نتیجه آنها را به خوبی توضیح نمی‌دهند، اثبات‌هایی که بر اساس چند خوش‌شانسی یا استدلال‌های اکتشافی که نتیجه می‌دهند ولی دقیق کردن آنها سخت است.

کار سختی است که نشان دهیم ترکیبیات اهدافی از نوعی که ذکر شد دارد (به استثنای مطلق مسئله $P = NP$) با این حال همان‌گونه که اهمیت یک نتیجه در ترکیبیات معمولاً خود نتیجه نیست، بلکه چیزی است غیرصریح‌تر که از اثبات یاد می‌گیریم. پس اهداف کلی ترکیبیات نیز همیشه به طور صریح بیان نمی‌شوند. برای روشن شدن این موضوع اجازه دهید به قضیه رمزی و کران تابع $R(k)$ برگردیم. علیرغم سادگی استدلال‌هایی که قبلاً ارائه کردم بهترین کران‌های شناخته شده را همین‌ها می‌دهند. اگر بخواهم دقیق‌تر بگویم مسئله زیر باز است.

مسئله. ۱) آیا ثابت $a > \sqrt{2}$ وجود دارد که $R(k) \geq a^k$ برای kهای به اندازه کافی بزرگ؟

۲) آیا ثابت $b < 4$ وجود دارد که برای kهای به اندازه کافی بزرگ داشته باشیم $R(k) \leq b^k$.

من این مسئله را یکی از مسائل مهم ترکیبیات می‌بینم و ماه‌های زیادی از عمر خود را بدون توفیق، برای حل آن تلاش کرده‌ام. ولی هنوز احساس خجالت می‌کنم که این را بنویسم، چرا که می‌دانم ریاضیدان‌های زیادی این سوال را بیشتر یک معما می‌بینند تا یک مسئله جدی ریاضی. علت ارادت زیاد من به این مسئله این است که برای من واضح شده است (همان‌گونه که به افراد زیاد دیگری که امتحانش کرده‌اند ثابت شده است) بسیار نامحتمل است این مسئله با یک استدلال هوشمندانه تک‌کاره که مخصوص این مسئله ساخته شده است حل شود (در واقع منظور من به طور خاص قسمتی از مسئله است که به کران بالای $R(k)$ مربوط می‌شود). اگر بخواهیم کمی در این مورد مبهم‌گویی کنیم اثبات حکم $R(k) \leq 4^k$ دارای روحیه‌ای موضعی است، به این مفهوم که بیشتر گراف را دور می‌ریزیم و روی همسایگی کوچکی از چند رأس تمرکز می‌کنیم.

به نظر می‌آید که کران بالای بهتر نیازمند استدلالی سرتاسری‌تر باشد، شامل کل گراف، و هیچ مدل کارایی برای چنین استدلالی در نظریه گراف موجود نیست. در نتیجه، حل این مسئله تقریبا مجبور است که تکنیک جدید بزرگی را در ترکیبیات ایجاد کند. یکی از دشواری‌های مسئله این است که، در حالی که رنگ‌آمیزی تصادفی کران پایین چندان بهتری از $2^{\frac{k}{2}}$ به دست نمی‌دهد، به نظر می‌آید که هیچ‌گونه فاصله گرفتنی از تصادفی بودن (مثل این که رأس‌ها را به پنج دسته تقسیم کنیم و احتمال این که یک یال قرمز باشد را برحسب این که در یک دسته قرار دارد و یا این که دو دسته را به هم وصل می‌کند تعیین کنیم) کران پایین را بهتر نمی‌کند. با این حال اگر کسی در این ایده پافشاری کند ، مجبور به تحلیل جداگانه‌ی اقسام مختلفی از گراف‌ها می‌شود که هیچ یک از آنها به خودی خود چالش برانگیز نیستند ولی پی‌گیری سیستماتیک آنها کار دشواری است. این من را به این خیال واداشت که ممکن است نوعی دسته‌بندی از رنگ‌آمیزی‌های آبی و قرمز (یا معادلاً از گراف‌ها) وجود داشته باشد. در این صورت با چک کردن این که در هر دسته کران کمتر از مثلا 3.99^k است می‌توانیم مسئله را حل کنیم. ایده‌ی رده‌بندی گراف‌ها در وهله‌ی اول عجیب به نظر می‌رسد، پس اجازه دهید نتیجه‌ای را که تا به حال بارها و بارها استفاده شده است را بیان کنم. ابتدا باید مفهومی از شبه تصادفی بودن را تعریف کنیم. فرض کنید G یک گراف است و A و B مجموعه‌های جدا از همی از رئوس G هستند. زوج (A,B) را ϵـ یکنواخت می‌گوییم اگر که عدد $\alpha > 0$ وجود داشته باشد که هرگاه $A' \subseteq A$ و $B' \subseteq B$ مجموعه‌هایی با عدد اصلی به ترتیب حداقل $\epsilon|A|$ و $\epsilon|B|$ باشند، تعداد یال‌هایی که A' را به B' وصل می‌کند با $\alpha|A'||B'|$ حداکثر به اندازه‌ی $\epsilon|A'||B'|$ اختلاف دارد. اگر ϵ کوچک باشد، این شرط به ما می‌گوید که گراف متشکل از یال‌های G که A را به B وصل می‌کند مانند یک گراف تصادفی با احتمال یال α است. نتیجه بعدی منسوب به زمردی [۱۴] است و به لم یکنواختی، یا لم نظم معروف است.

قضیه. فرض کنید G یک گراف است و ϵ عدد مثبتی است و k یک عدد طبیعی است. در این صورت ثابت K وجود دارد که فقط به k و ϵ بستگی دارد، به طوری که رأس‌های G را می‌توان به m زیرمجموعه‌ی $A_1, A_2, \cdots, A_m$ افراز کرد که در آن $k \leq m \leq K$، به طوری که حداقل $(1-\epsilon)\binom{m}{2}$ از زوج‌های (A_i, A_j) (که $i < j$ است) ϵـ یکنواخت باشند.

این قضیه، همان‌طور که خواننده ی هوشیار متوجه شده است، بیان دقیقی از اصل کلی (۲) است که قبلاً بیان کردیم (ولی تأکید می‌کنم که همه‌ی این اصول را نمی‌توان دقیق کرد). متاسفانه با این که لم یکنواخت‌سازی زمردی ابزاری ایده‌آل برای مسائل پرشماراست، مسائل دیگری، مانند کران پایین بهتر برای قضیه رمزی هستند، که این لم در مورد آنها چیزی برای گفتن ندارد. پس یکی از اهداف کلی نظریه گراف این است که رده‌بندی‌های مفصل‌تر و ظریف‌تری را بیابد و هدفی دیگر، که به گونه‌ای مخالف قبلی است، این است که راه‌حل‌هایی را بدون استفاده از لم یکنواختی زمردی پیدا کند. پیشرفت قابل توجه در هر یک از این دو پروژه تأثیر متقابل قابل توجهی را روی نظریه گراف خواهد داشت (این به گونه‌ای حشو است چرا که یک معیار خوب برای میزان پیشرفت این است که آیا اجازه حل سوالات جالب را می‌دهند؟). به طور کلی با کسب مهارت در حل مسائل در زمینه‌ای مانند ترکیبیات متوجه خواهیم شد که برخی دشواری‌ها مرتباً تکرار می‌شوند. ممکن است نتوانیم این دشواری‌ها را در قالب یک حدس به خوبی بیان کنیم، پس به جای آن روی مسئله خاصی متمرکز می‌شویم که شامل این دشواری‌ها باشد. در این حالت این مسئله اهمیتی می‌یابد که فراتر از این است که جواب مسئله آری یا نه است. این موضوع روشن می‌کند که چگونه تعداد زیادی از مسائل اردوش دارای عمق پنهان زیادی بوده‌اند. اما تکلیف این انتقاد که ترکیبیات زمینه‌ای سطحی است چه می‌شود؟ یکی از خرسندی‌های ریاضیات این است که به قول معروف، با سوار بودن بر دوش غول‌ها، می‌توانیم به نتایج مرتفعی برسیم که برای نسل‌های قبلی قابل تصور نبوده است. حال آن که، بیشتر مقالات ترکیبیات همه‌ی آنچه نیاز است را در خود دارند، یا این که پیشینه اطلاعاتی کمی را از خواننده مطالبه می‌کنند. این را با مقاله‌ای در نظریه جبری اعداد مقایسه کنید که اگر با اطلاعات معمولی دروس کارشناسی شروع کنیم، درک آن ممکن است سال

ها طول بکشد.

این انتقاد تفاوت اولویت‌های نظریه‌سازان و مسئله‌حل‌کن‌ها را منعکس می‌کند. یک نظریه‌ساز تمایل دارد که بگوید قضیه A عمیق است چرا که از قضیه B استفاده می‌کند که آن هم از قضیه C استفاده می‌کند و قس علی هذا، که همه‌ی اینها به طور منفرد نتایج قابل توجهی هستند. یک مسئله‌حل‌کن ممکن است که چنین زنجیری طولانی از وابستگی‌های منطقی را در ذهن خود نداشته باشد. با این حال، اگر که ما نوع مناسب‌تری از وابستگی را، دوباره بر اساس اصول کلی، در نظر بگیریم در این صورت ماجرا متفاوت می‌شود. اکثریت مواقع این‌گونه خواهد بود که، در حالی که هیچ وابستگی صوری بین دو نتیجه وجود ندارد، هیچ امیدی برای حل یکی از آنها وجود نخواهد داشت اگر که ما از اصول کلی معرفی شده در اثبات دیگری مطلع نباشیم. زنجیرهای وابستگی از این نوع می‌توانند بسیار طویل باشند، پس ترکیبیات‌دان‌ها نیز می‌توانند از این موضوع خشنود باشند که مسائلی را حل کرده‌اند که بسیار دور از دسترس نسل پیش بوده‌اند. در این حالت فرد احساس می‌کند که موضوع در کلیت خود در حال پیشرفت است.

تا بحال تمامی استدلال‌های ما مربوط به درون ترکیبیات بوده است. سعی کرده‌ام که نشان دهم که این موضوع دارای انسجام و جهت‌گیری است که برای بیگانگان واضح نیست ولی با این همه، مهم است. با این حال، چیزی در مورد این که ریاضیات در کلیت خود چگونه می‌تواند از پیشرفت ترکیبیات بهره جوید نگفته‌ام. اجازه دهید ببینیم که اتیه در این رابطه چه می‌گوید [۳]:

... توجیه نهایی برای انجام دادن ریاضی در ارتباط تنگاتنگ با یکپارچگی کلی آن است. اگر ما قبول کنیم که، از دیدی کاملاً کارکردگرایانه ریاضیات با استفاده از کاربردهایش خودش را توجیه می‌کند، در این صورت کل ریاضیات موجه خواهد شد، در صورتی که یک کل به هم پیوسته باقی بماند. هر بخشی از آن که از تنه‌ی اصلی جدا شود، باید خود را به گونه‌ی مستقیم‌تری توجیه کند.

اگر نیاز به چنین توجیهی را بپذیریم ترکیبیات چه چیزی برای گفتن دارد؟ یک نکته‌ی روشن این است که می‌توانیم ترکیبیات را به طور مستقیم توجیه کنیم، به خاطر ارتباط نزدیک آن به علوم کامپیوتر، که کارکردهای آن مشخص هستند (عجیب این که وقتی کمیته برنامه‌ریزی کنگره بین‌المللی ریاضیدان‌ها در سال ۱۹۹۸ ارتباط بین قسمت‌های مختلف ریاضی را لیست می‌کردند، این ارتباط را تشخیص ندادند) و اما ارتباط با دیگر زمینه‌ها، کاربردهایی از ترکیبیات در زمینه‌های احتمالات، نظریه مجموعه‌ها، رمز نگاری، نظریه ارتباطات، هندسه فضاهای باناخ، آنالیز هارمونیک، نظریه اعداد و ... وجود دارد. با این وجود، الآن که این را می‌نویسم می‌دانم که خیلی از این کاربردها نمی‌توانند توجه مثلاً یک هندسه دیفرانسیل‌دان را جلب کند، او ممکن است همه‌ی این‌ها را به‌گونه‌ای متعلق به آن قسمت دور و پرفاصله از ریاضیات ببیند که می‌توان بدون خطر آن را نادیده گرفت. حتی کاربردها در نظریه اعداد "نوع بدی" از نظریه اعداد هستند. شاید مفید باشد که راه‌های مختلفی که یک شاخه از ریاضیات می‌تواند برای بقیه مفید باشد را بررسی کنیم. در اینجا لیستی را که بر اساس مستقیم بودن مرتب شده را آورده‌ایم که چگونه زمینه A میتواند به زمینه B کمک کند. ۱)قضیه‌ای از A بلافاصله منجر به نتیجه‌ای مفید در B می‌شود.

۲)قضیه‌ای از A نتیجه‌ای در B دارد ولی اثبات آن نیازمند مقداری کار است.

۳)قضیه‌ای در A به مسئله‌ای در B به‌قدری شبیه است که ما را قادر می‌سازد از اثبات A تقلید یا اقتباس کنیم و مسئله را در B پاسخ بدهیم.

۴)برای این که مسئله‌ای در A را حل کنیم، انگیزه پیدا می‌کنیم که ابزارهایی در B را درست کنیم که دارای جذابیت مستقل هستند.

۵)زمینه B شامل تعریفاتی است که به تعریف های A شبیه هستند. (اگر بخواهم یک مثال ارائه کنم، برخی اوقات فرد می‌خواهد که مفهومی از استقلال را تعریف کند که از بعضی جهت ها ،ولی نه به‌طور کامل،مثل استقلال در فضاهای برداری رفتار کند). در این حالت زمینه A روش‌های پرباری را برای دسته‌بندی و فکر کردن روی نتایج و مسائل B پیشنهاد می‌دهد.

۶)اگر کسی در زمینه A متخصص شود، در این صورت فرد عادت‌هایی فکری را کسب می‌کند که او را قادر می‌سازد کمک‌های قابل توجهی را به زمینه B ارائه کند.

۷)زمینه A در روح خود به زمینه B نزدیک است، به طوری که کسی که در زمینه A مستعد باشد، احتمالاً در زمینه B نیز مستعد است. درضمن ریاضیدان‌های زیادی در هر دو زمینه کار می‌کنند. ارتباطات غیرمستقیم‌تر مثل ۴ تا ۷ به همان نسبت هم غیرآشکارتر هستند. با این همه سهم آنها در یک‌پارچگی ریاضیات را نباید دست کم گرفت. من خود بالاخص این احساس را بعد از سال‌ها کار کردن در هندسه فضاهای باناخ دارم. علت اولیه من برای کار کردن در این زمینه این بود که نتایجی مانند قضیه دورتسکی و مسائل باز طبیعی زیادی را بسیار جالب می‌دیدم. بعداً متوجه شدم که زمینه منتخب من قویاً به خاطر جدا شدن از ریشه‌های اصلی‌اش در معادلات دیفرانسیل مورد انتقاد است و در پی آن هدفش را گم کرده است. من تصدیق می‌کنم که ارتباطات زیادی از نوع ۱ و ۲ از نظریه محض فضاهای باناخ به بقیه زمینه‌ها وجود ندارد، هرچند که ایده‌های عمیقی وجود دارند که عده زیادی معتقد هستند روی آنها به اندازه کافی کار نشده است. ولی، به محض این که ارتباطات ضعیف‌تر را در نظر بگیریم به وفور شاهد آنها خواهیم بود. استخراج تمرکز اندازه مثال خوبی از ۳ است (یا ۴ ــ دسته‌بندی تا حدودی مصنوعی است.) و برای ۶ و ۷، می‌توانم از تجربه شخصی خود بگویم، اخیراً روی مسائل زیادی خارج از نظریه فضاهای باناخ کار کرده‌ام؛ با این که نتایج فضاهای باناخ را استفاده نکرده‌ام، تجربه‌ی من در این زمینه من را قادر ساخت که به مسائل به گونه‌ای فکر کنم که در حالت دیگری به ذهنم نمی‌رسید. به هیچ وجه این گونه نیست که من در این مورد تنها باشم؛ بسیاری از ریاضیدان‌هایی که در زمینه فضاهای باناخ کار کرده بودند، در زمینه‌های دیگری مانند آنالیز هارمونیک ،معادلات دیفرانسیل با مشتقات جزئی ،جبرهای C* ، احتمالات و ترکیبیات کارهای موفقی داشته‌اند. تنها راهی که می‌توان این ارتباطات را نادیده گرفت این است که فرد باور داشته باشد که همه‌ی این ارتباطات بین دو زمینه کم ارزش و غیرجالب است. این دیدگاهی افراطی است و قسمت زیادی از ریاضیات نوین شامل نتایج زیادی با کاربردهای عملی مستقیم را معزول می‌کند، و دشوار است که باور کنیم کسی این دیدگاه را جدی می‌گیرد. با این حال به نظر می‌رسد بعضی این دید را دارند که باز این ایده را به ذهن می‌رساند که دو فرهنگ در ریاضیات محض داریم. احتمالاً این حرف درستی است که ارتباطات بیشتری بین قسمت‌های مسئله‌حل‌کردنی و نظریه‌سازی ریاضی وجود دارد تا این که بین آنها، و همین علت آن است که برچسب "دو فرهنگ" مناسب است (باید دوباره بگویم که این نام‌گذاری‌ها فراساده‌سازی هستند و من کاملاً متوجه هستم که افراد زیادی به آنچه من نظریه سازی نامیده ام از این جهت جذب شده‌اند که می‌خواستند مسائل خاصی را پاسخ دهند). اگر درست باشد که ریاضی محض به دو فرهنگ گسترده تقسیم می‌شود که ارتباط چندانی بین آنها وجود ندارد، هنوز می‌توان پرسید که آیا این موضوع حائز اهمیت است؟ به نظر من بله. یک دلیل این است که این وضعیت نتایج عملی نامطلوبی را دارد. برای مثال ریاضیدانان از یک فرهنگ ممکن است تصمیم‌هایی بگیرند که آینده‌ی کاری ریاضیدان‌های فرهنگ دیگر را تحت تأثیر قرار دهد. اگر که درک متقابل کمی بین این دو گروه وجود داشته باشد، تصمیم‌گیری عادلانه، که در بهترین شرایط، کار سختی است، سخت‌تر می‌شود. (من به شخصه گله‌مند نیستم، ولی من در وضعیت کاری خود خوش‌شانس بوده‌ام.) اثر دوم این است که دانشجویان پژوهشگری که به طور طبیعی مناسب یک فرهنگ هستند ممکن است خود را تحت این فشار بیابند که در زمینه‌ای از فرهنگ دیگر کار کنند و در نهایت استعداد خود را تلف کنند. این خطری شایع برای دانشکده‌ای است که تحت سلطه‌ی تعداد کمی موضوع قرار گرفته است. این تأثیرات شاید محصولات جانبی زندگی آکادمیک باشند، و دلایل اصلی من برای مصر بودن به ارتباط بهتر بین این دو گروه نیست. مهم ترین اشکال کمبود ارتباط این است که نشان از یک فرصت بزرگ از دست رفته است. من بعضاً از ریاضیدانانی از جناح نظریه‌سازان گلایه از مسئله‌ای را شنیده‌ام که همه‌ی ابزارهای موجود را روی آن امتحان کرده‌اند ولی یک هسته‌ی سرسخت باقی‌مانده است که "دراصل ترکیبیات" است. این راهی است برای گفتن این که مسئله خیلی سخت است، ولی، به عبارتی دقیق‌تر، به گونه‌ای سخت است که دقیقاً ریاضی‌دان‌های با ذهن مسئله‌حل‌کن را جذب می‌کند. متأسفانه همچنین سخت است که به درجه‌ای از فهم برسیم که ذات در اصل ترکیبیات مسئله را درک کنیم

. چنین شرایطی مخصوص همکاری بین فرهنگی است، ولی چنین همکاری نیازمند تلاش زیادی از طرف مسئله‌حل‌کن‌ها است که کمی تئوری یاد بگیرند، و احساس همدردی بیشتری از طرف نظریه‌سازان نسبت به ریاضیدان‌هایی که نمی‌دانند کوهومولوژی چیست، است. این تلاش‌ها قطعاً مایه غنای هر دو فرهنگ ریاضی خواهند بود.

مراجع

[1] M. F. Atiyah. How research is carried out. Bull. I.M.A. 10, pages 232–4, 1974.

[2] M. F. Atiyah. An interview with michael atiyah. Math. Intelligencer 6, pages 9–19, 1984.

[3] M. F. Atiyah. Identifying progress in mathematics. ESF conference in Colmar, C.U.P., pages 24–41, 1985.

[4] A. Dvoretzky. Some results on convex bodies and banach spaces. Proc. Symp. on Linear Spaces, Jerusalem, pages 123–160, 1961.

[5] P. Erdos. Some remarks on the theory of graphs. Bull. A.M.S. 53, pages 292–4, 1987.

[6] P. Erdos and L. Lovasz. Problems and results on 3-chromatic hypergraphs and some related questions. in A. Hajnal et al. eds, Infinite and finite sets, North-Holland, Amsterdam, pages 609–628.

[7] R. L. Graham F. Chung and R. M. Wilson. Quasi-random graphs. Combinatorica 9, pages 345–362, 1990.

[8] W. T. Gowers. The two cultures of mathematics. Mathematics: Frontiers and Perspectives, pages 65–79, 2000.

[9] S. Janson. Poisson approximation for large deviations. Random structures and Algorithms 1, pages 221–230.

[10] B. S. Kashin. Sections of some finite dimensional sets and classes of smooth functions. Isv. ANSSSR, ser. mat. 41 (Russian), pages 334–351, 1977.

[11] V. D. Milman. A new proof of the theorem of A. Dvoretzky on sections of convex bodies. Funct. Anal. Appl. 5 (translated from Russian), pages 28–37, 1971.

[12] K. Roth. On certain sets of integers. J. London Math. Soc. 28, pages 245–252, 1953.

[13] S. J. Szarek. On Kashin's almost Euclidean orthogonal decomposition of l_1^n. Bull. Acad. Polon. Sci. 26, pages 691–694, 1978.

[14] E. Szemeredi. Regular partitions of graphs. in Proc. Colloque Inter. CNRS, J.-C. Bermond et al. eds, pages 399–401, 1978.

[15] A. Thomason. Pseudo-random graphs. in M. Karonski, ed., Proceedings of Random Graphs, Poznan, Annals of Discrete Mathematics 33, pages 307–331, 1985.

چند مساله

تینا ترکمان، علی چراغی

مساله ها

مساله ۱. فرض کنید p یک چندجمله‌ای تکین با ضرایب مختلط باشد. ثابت کنید z_0 ای روی دایره ی واحد هست به‌طوری که

$$|p(z_0)| \geq 1$$

مساله ۲. فرض کنید A و B دو ماتریس مربعی روی میدان $\mathbb{R}$ باشند که با هم جابجا می‌شوند. فرض کنید $\det(A+B) \geq 0$. ثابت کنید برای هر $n \in \mathbb{N}$ داریم

$$\det(A^n + B^n) \geq 0$$

مساله ۳. برای یک چندجمله‌ای

$$p(x) = a_m x^m + a_{m-1}x^{m-1} + \cdots + a_0$$

با ضرایب حقیقی قرار دهید

$$\Gamma(p) = a_m^2 + a_{m-1}^2 + \cdots + a_0^2$$

فرض کنید $f(x) = 3x^2 + 7x + 2$ و $g(x)$ را به گونه‌ای پیدا کنید که دو شرط زیر برقرار باشد،

۱) $g(0) = 1$

۲) برای هر $n \geq 1$، $\Gamma(f^n) = \Gamma(g^n)$ که در این جا منظور از f^n و g^n به توان n ام رساندن است.

مساله ۴. آیا دامنه ی صحیح D که $D \subsetneqq \mathbb{R}$ وجود دارد که $\mathbb{R}$ میدان کسرهای آن باشد؟

مساله ۵. آیا یک مجموعه‌ی نامتناهی شمارا وجود دارد که یک خانواده‌ی ناشمارا از زیرمجموعه‌های آن باشد به طوری که اشتراک هر دو تای متمایز از آنها متناهی باشد؟

مساله ۶. فرض کنید H یک زیرگروه متناهی از گروه توابع پیوسته و یک به یک و پوشا از بازه‌ی $[0,1]$ به خودش باشد. نشان دهید $|H| \leq 2$.

مساله ۷. فرض کنید k و n اعدادی طبیعی و مثبت باشند و $k > 1$. R را حلقه‌ای در نظر بگیرید که لزوما یکدار نیست و در دو شرط زیر صدق می کند،

۱) R دارای حداقل یک عضو غیر پوچ‌توان است.

۲) اگر $x_1, x_2, \cdots, x_k$ اعضای ناصفری از حلقه باشند، آنگاه

$$x_1^n + x_2^n + \cdots + x_k^n = 0$$

نشان دهید که R یک حلقه‌ی تقسیم است.

مساله ۸. فرض کنید $C = C^0(\mathbb{R})$ حلقه‌ی توابع پیوسته روی $\mathbb{R}$ باشد و D حلقه‌ی توابع مشتق‌پذیر روی $\mathbb{R}$ باشد. ثابت کنید D زیرحلقه‌ای یکریخت با C ندارد که شامل عنصر همانی حلقه (تابع ثابت ۱) باشد. جمع و ضرب را همان جمع و ضرب مقدار توابع در هر نقطه در نظر می‌گیریم.

مساله ۹. فرض کنید $f : [0,1] \to [0,\infty)$ تابعی پیوسته باشد به طوری که $f(0) = f(1) = 0$ و $f(x) > 0$ برای هر $x \in (0,1)$. ثابت کنید مربعی وجود دارد که دو رأسش روی محور x و دو رأس دیگرش روی نمودار f قرار دارند.

مساله ۱۰. فرض کنید H ماتریسی حقیقی، مربعی و متقارن باشد که مقادیر ویژه‌ی متمایز داشته باشد و A ماتریسی حقیقی با ابعاد برابر با H باشد. فرض کنید

$$H_0 = H, H_1 = AH_0 - H_0A, H_2 = AH_1 - H_1A$$

متقارن باشند. ثابت کنید $AA^T = A^TA$.

پاسخ ها

پاسخ ۱. فرض کنید برای همه‌ی zهای روی دایره ی واحد داشته باشیم $|p(z)| < 1$ و قرار دهید $n = \deg p$ و فرض کنید $n \geq 1$، زیرا برای $n = 0$ حکم بدیهی است. حال قرار دهید $f(z) = -z^n$ و $g = p$. در این صورت برای هر z روی دایره ی واحد داریم،

$$|g(z)| = |p(z)| < 1 = |-z^n| = |f(z)|$$

پس روی دایره ی واحد داریم $|g| < |f|$ و طبق قضیه روشه[1] تعداد ریشه‌های با احتساب تکرر f و $f+g$ درون دایره ی واحد با هم یکسان است. اما f دقیقا n ریشه درون دایره دارد و $f+g$ حداکثر $n-1$ ریشه دارد زیرا یک چندجمله‌ای ناصفر با درجه ی حداکثر $n-1$ است. این تناقض کار را تمام می‌کند.

پاسخ ۲. داریم

$$A^n + B^n = \prod_{\zeta}(A - B\zeta)$$

که ضرب روی ریشه‌های چند جمله ای $x^n + 1$ محاسبه می شود. ریشه‌های حقیقی x^n+1 در حالتی که n فرد است، فقط -1 است و در حالت n زوج ریشه حقیقی وجود ندارد. پس در هر حال می‌توان این ضرب را به شکل

$$A^n + B^n = (A+B)^\epsilon \prod_{\zeta}(A - B\zeta)(A - B\bar{\zeta})$$

نمایش داد که البته این‌بار ضرب روی ریشه‌های غیر حقیقی $x^n + 1$ محاسبه می شود و از بین هر زوج ریشه ی مزدوج مختلط تنها یکی را در نظر می گیریم و ϵ صفر یا یک است بر حسب این که n زوج یا فرد باشد. پس داریم

$$\det(A^n+B^n) = \det(A+B)^\epsilon \prod_{\zeta} \det(A-B\zeta)(A-B\bar{\zeta})$$

پس کافی است برای هر ماتریس M اثبات کنیم $\det M\overline{M} \geq 0$ که این هم واضح است زیرا،

$$\det M\overline{M} = \det M \det \overline{M} =$$
$$\det M.\overline{\det M} = |\det M|^2 \geq 0$$

پاسخ ۳. تعریف کنید

$$\gamma\Big(p(x)\Big) = p(x)p(\frac{1}{x})$$

در این صورت $\Gamma(p)$ ضریب جمله ی ثابت در چندجمله‌ای لوران[2] بالا خواهد بود. حال از آنجا که $\gamma\Big(p(x)\Big)$ نسبت به تعویض x با $\frac{1}{x}$ تغییر نمی‌کند و ضربی است و $\gamma\Big(x^n\Big) = 1$ داریم،

$$\begin{aligned}\gamma\Big(f(x)^n\Big) = \ & \gamma\Big((3x+1)^n(x+2)^n\Big) = \\ & \gamma\Big((3x+1)^n\Big)\gamma\Big((x+2)^n\Big) = \\ & \gamma\Big((3x+1)^n\Big)\gamma\Big((\frac{1}{x}+2)^n\Big) = \\ & \gamma\Big((3x+1)^n\Big)\gamma\Big((1+2x)^n\Big) = \\ & \gamma\Big(((3x+1)(1+2x))^n\Big) = \\ & \gamma\Big((6x^2+5x+1)^n\Big)\end{aligned}$$

پس کافی است قرار دهیم $g(x) = 6x^2 + 5x + 1$.

[1] Rouché
[2] Laurent

پاسخ ۴. بله وجود دارد. ابتدا یک مجموعه ی مستقل جبری روی $\mathbb{Q}$ از اعداد حقیقی پیدا کنید که ماکسیمال باشد. منظور از مجموعه ی مستقل جبری روی $\mathbb{Q}$، مجموعه ای از اعداد است که هیچ چند جمله ای ناصفر چند متغیره با ضرایب گویا وجود نداشته باشد که اگر اعضای آن مجموعه را به جای متغیرهای ورودی چندجمله ای قرار دهیم برابر با صفر شود. ساختن این مجموعه ی ماکسیمال به وسیله ی لم زرن[3] قابل انجام است. این مجموعه را T بنامید و توجه کنید اگر $x \in \mathbb{R}\backslash\mathbb{Q}[T]$ آن گاه x روی $\mathbb{Q}[T]$ جبری است، یعنی ریشه ی یک چند جمله ای ناصفر با ضرایب داخل $\mathbb{Q}[T]$ است، که در این جا $\mathbb{Q}[T]$ حلقه ی تشکیل شده توسط اعداد گویا و اعضای T است. دلیلش این است که با افزودن x به T به یک مجموعه ی وابسته ی جبری می رسیم پس چند جمله ای نابدیهی ای شامل x صفر می شود. اکنون فرض کنید D بستار صحیح[4] حلقه ی $\mathbb{Q}[T]$ باشد، یعنی مجموعه ی تمام اعداد حقیقی مانند x که به ازای یک $n \geq 1, n \in \mathbb{N}$ و اعدادی مانند $a_i \in \mathbb{Q}[T]$ داشته باشیم

$$x^n + a_{n-1}x^{n-1} + ... + a_0 = 0$$

ادعا می کنیم D دامنه ی صحیح مورد نظر است. ابتدا توجه کنید D شامل $\mathbb{Q}[T]$ و در نتیجه شامل $\mathbb{Q}$ هست. از طرفی D یک حلقه است زیرا اگر α, β اعدادی حقیقی در D باشند، این اعداد ریشه ی چندجمله ای هایی تکین با ضرایب داخل $\mathbb{Q}[T]$ هستند، و به کمک قضیه ی اساسی چند جمله ای های متقارن[5] می توان ثابت کرد که اعداد $\alpha\beta$ و $\alpha + \beta$ و $-\alpha$ نیز چنین خاصیتی دارند. اثبات مشابه حکم معروفی است که در آن نشان می دهیم اگر α, β روی میدان F جبری باشند، $\alpha\beta$ و $\alpha + \beta$ نیز جبری هستند، تنها تفاوت این است که در این جا ضرایب از یک دامنه ی صحیح می آیند. پس D یک حلقه ی جابجایی و یکدار است و چون زیر حلقه ی $\mathbb{R}$ است پس دامنه ی صحیح است. از طرفی D میدان نیست، زیرا اگر $t \in T$ عضوی از مجموعه ی مستقل جبری فوق باشد، ادعا می کنیم $\frac{1}{t}$ در D قرار ندارد. اگر این طور نباشد پس عدد طبیعی $n \geq 1$ و ضرایب $a_i \in \mathbb{Q}[T]$ وجود دارند که $(\frac{1}{t})^n + a_{n-1}(\frac{1}{t})^{n-1} + ... + a_0 = 0$ پس داریم،

$$1 + a_{n-1}t + a_{n-2}t^2 + ... + a_0t^n = 0$$

اما همگی ضرایب و همچنین t داخل $\mathbb{Q}[T]$ هستند. این خود یک چند جمله ای نابدیهی است زیرا جمله ی ثابت آن برابر با 1 است و روی اعضای T صفر شده است و این با انتخاب T در تناقض است. پس D میدان نیست و در نتیجه $D \subsetneq \mathbb{R}$. در انتها ثابت می کنیم میدان کسر های D کل $\mathbb{R}$ است. فرض کنید $x \in \mathbb{R}$ است. اگر x داخل T باشد که حکم واضح است و گرنه x روی $\mathbb{Q}[T]$ جبری است و داریم

$$a_nx^n + a_{n-1}x^{n-1} + ... + a_0 = 0$$

که $a_i \in \mathbb{Q}[T]$ و $a_n \neq 0$. اکنون می توانیم محاسبات زیر را روی میدان کسر های D انجام دهیم،

$$\begin{aligned} 0 &= a_nx^n + a_{n-1}x^{n-1} + ... + a_0 \\ &= a_n\Big(\frac{a_nx}{a_n}\Big)^n + a_{n-1}\Big(\frac{a_nx}{a_n}\Big)^{n-1} + ... + a_0 \\ &\Rightarrow a_n(a_nx)^n + a_{n-1}a_n(a_nx)^{n-1} + ... + a_0a_n^n = 0 \\ &\Rightarrow (a_nx)^n + a_{n-1}(a_nx)^{n-1} + ... + a_0a_n^{n-1} = 0 \end{aligned}$$

که در آخرین گام از دامنه ی صحیح بودن D استفاده کردیم و a_n را حذف کردیم. در آخرین رابطه یک چند جمله ای تکین با ضرایب داخل $\mathbb{Q}[T]$ در مقدار $y = a_nx$ صفر شده است. پس $y \in D$. از طرفی به وضوح $a_n \in \mathbb{Q}[T] \subset D$ پس $x = \frac{y}{a_n} \in \mathrm{Frac}(D)$ یعنی x در میدان کسر های D است. پس میدان کسر های D برابر با $\mathbb{R}$ است.

پاسخ ۵. بله وجود دارد. اعداد گویا را در نظر بگیرید و برای هر $\alpha \in \mathbb{R}$ یک دنباله از اعداد گویا انتخاب کنید که به α میل کند و اعضای این دنباله را در یک مجموعه‌ی S_α قرار دهید. در این صورت به دلیل یکتایی حد این خانواده از زیرمجموعه‌های اعداد گویا خواص موردنظر را دارد.

[3] Zorn
[4] Integral closure
[5] The fundamental theorem of symmetric polynomials

پاسخ ۶. فرض کنید g عضوی از H باشد. چون g تابعی پیوسته و یک به یک است پس به طور اکید یکنواست. از پوشایی آن نتیجه می گیریم یا $g(1)=0$ و $g(0)=1$ یا $g(0)=0$ و $g(1)=1$. حال اگر g اکیدا صعودی باشد و به ازای a ای $g(a)<a$ آنگاه برای هر $n\geq 1$ به استقرا نتیجه می گیریم $g^n(a)<g^{n-1}(a)$ و $g^n(a)<a$. اما چون g عضوی از گروه متناهی H است پس طبق قضیه لاگرانژ[6]، $g^{|H|}=1$ پس $g^{|H|}$ تابع همانی است و $g^{|H|}(a)=a$ که تناقض است. حالتی که g اکیدا صعودی باشد و به ازای a ای $g(a)>a$ نیز به طور مشابهی رد می شود. پس برای هر $a\in[0,1]$، $g(a)=a$. یعنی تنها عضو اکیدا صعودی در H همان تابع همانی است. حال توجه کنید اگر g و h اکیدا نزولی باشند آنگاه $g\circ h$ اکیدا صعودی و عضوی از گروه است، پس همانی است و h وارون g است و در اصل برابر با g است. پس مرتبه ی H برابر با 1 یا 2 است.

پاسخ ۷. فرض کنید a یک عضو غیر پوچ‌توان و x یک عضو ناصفر در R باشند. طبق شرط دوم، $kx^n=0$ و $(k-1)x^n+a^n=0$. پس اگر $a^n=e$، برای هر x ناصفر داریم $x^n=e$، دقت کنید تا این جا فرض نکردیم که e عنصر همانی ضرب است. از این نتیجه می‌شود هیچ عنصر ناصفر پوچ‌توانی در حلقه نیست. هم‌چنین داریم $e^2=(a^2)^n=e$. ادعا می‌کنیم e همانی ضرب در حلقه است. ابتدا دقت کنید برای هر x ناصفر، $ex=x^nx=xx^n=xe$ و البته این رابطه برای $x=0$ هم درست است. حال که می‌دانیم x و e جابه‌جا می‌شوند، نتیجه می گیریم،

$$(x-ex)^n=\sum_{i=0}^{n}(-1)^{n-i}\binom{n}{i}x^i(ex)^{n-i}=$$

$$x^n+e\Big(\sum_{i=0}^{n}(-1)^{n-i}\binom{n}{i}x^n\Big)-ex^n=$$

$$x^n+e(x-x)^n-ex^n=x^n-ex^n=0$$

پس $x-ex=0\Rightarrow x=ex=xe$، و e همانی حلقه است. پس برای هر x ناصفر عنصر وارون x^{n-1} وجود دارد. پس R حلقه ی تقسیم است.
نکته: چون $x^{n+1}=x$ از قضیه‌ی معروفی از جیکوبسن[7] نتیجه می‌شود حلقه جابه‌جایی است. پس R میدان است.

پاسخ ۸. نشان می‌دهیم هر همریختی[8] حلقه ای از C به D هر تابع را به تابع ثابت تصویر می‌کند. این نشان خواهد داد C با هیچ زیر حلقه ی D یکریخت نیست وگرنه می شد همریختی ای بین C و D یافت که بین C و آن زیرحلقه، یکریختی[9] باشد. فرض کنید T یک همریختی از C به D باشد. هر تابع f در C را می‌توان به طور یکتا به صورت f_+-f_- نوشت که $f_+\geq 0$ و $f_-\geq 0$ و $f_+f_-=0$، کافی است قرار دهید $f_+=\frac{|f|+f}{2}$ و $f_-=\frac{|f|-f}{2}$. در این صورت چون

$$T(f_+)-T(f_-)=T(f_+-f_-)=T(f),$$
$$T(f_+)T(f_-)=T(f_+f_-)=0,$$
$$T(f_+)=T(\sqrt{f_+}.\sqrt{f_+})=T(\sqrt{f_+})T(\sqrt{f_+})\geq 0,$$
$$T(f_-)=T(\sqrt{f_-}.\sqrt{f_-})=T(\sqrt{f_-})T(\sqrt{f_-})\geq 0$$

پس $T(f_+)=T(f)_+$ و $T(f_-)=T(f)_-$. دقت کنید که تابع ثابت 1 باید به خودش برود. قرار دهید $g=T(f)$ و فرض کنید برای یک t_0 رابطه ی $g'(t_0)\neq 0$ برقرار باشد. چون $T(f-g(t_0))=g-g(t_0)$ داریم

$$\Big(g-g(t_0)\Big)_+=T\Big(\big(f-g(t_0)\big)_+\Big)\in D$$

از آن جا که $g'(t_0)\neq 0$ تابع $g-g(t_0)$ در $t=t_0$ تغییر علامت می‌دهد. پس یکی از $\lim_{t\to t_0^+}\frac{(g(t)-g(t_0))_+}{t-t_0}$ و $\lim_{t\to t_0^-}\frac{(g(t)-g(t_0))_+}{t-t_0}$ برابر با 0 و دیگری برابر با $g'(t_0)$ است. چون $g'(t_0)\neq 0$ با مشتق پذیری $\big(g-g(t_0)\big)_+$ به تناقض می رسیم. این تناقض نشان می دهد برای هر t، $g'(t)=0$ پس g تابع ثابت است و ادعای ما اثبات شد.

پاسخ ۹. تابع f را با برابر 0 قرار دادن در نقاط تعریف

[6] Lagrange
[7] Jacobson
[8] Homomorphism
[9] Isomorphism

نشده به روی $[0,\infty)$ گسترش می‌دهیم. اکنون تابع g را به صورت $g(x)=f\big(x+f(x)\big)-f(x)$ در نظر می‌گیریم. فرض کنید برای یک $m\in(0,1)$ مقدار $f(m)$ ماکسیمم باشد، دقت کنید ماکسیمم حتما درون بازه اتفاق می افتد چون در دو سر بازه مقدار f صفر است. این ماکسیمم تابع روی کل $\mathbb{R}$ نیز هست و داریم $g(m)\leq 0$. از طرفی چون

$$0+f(0)=0<m$$
$$m+f(m)>m$$

پس طبق قضیه مقدار میانی y ای در بازه $(0,m)$ وجود دارد که $m=y+f(y)$. پس $g(y)\geq 0$. در نتیجه طبق قضیه مقدار میانی عددی بین m و y وجود دارد که g در آن صفر می‌شود. اکنون دیگر به وضوح می توانید رأس های مربع را بیابید،

$$\Big(x,0\Big),\Big(x+f(x),0\Big),\Big(x,f(x)\Big),\Big(x+f(x),f\big(x+f(x)\big)\Big)$$

پاسخ ۱۰. با تغییر پایه ی همزمان همه ی ماتریس ها توسط یک ماتریس متعامد، فرض و حکم عوض نمی‌شوند. چون H_0 حقیقی و متقارن است، با تغییر پایه می‌توان فرض کرد که H_0 قطری است و روی قطرش مولفه‌های متمایز دارد. چون H_1 متقارن است نتیجه می‌شود که

$$AH_0-H_0A=(AH_0-H_0A)^T=H_0A^T-A^TH_0$$

یعنی $(A+A^T)H_0=H_0(A+A^T)$. پس به دلیل متمایز بودن مقدار ویژه های روی قطر H_0 ماتریس $A+A^T$ نیز قطری است. قرار دهید $D=\frac{1}{2}(A+A^T)$، $S=\frac{1}{2}(A-A^T)$. حال $A=D+S$ و D متقارن است. چون H_2 متقارن است داریم $(A+A^T)H_1=H_1(A+A^T)$ و $H_1D=DH_1$. چون D و H_0 جابه‌جا می‌شوند، $AH_0-H_0A=SH_0-H_0S$ و

$$D(SH_0-H_0S)=D(AH_0-H_0A)=DH_1=$$
$$H_1D=(AH_0-H_0A)D=(SH_0-H_0S)D$$

پس خواهیم داشت $H_0(DS-SD)=(DS-SD)H_0$. از متمایز بودن مقدار ویژه های H_0 نتیجه می گیریم $DS-SD$ نیز قطری است. چون D قطری است داریم

$$(DS-SD)_{i,i}=(DS)_{i,i}-(SD)_{i,i}=$$
$$D_{i,i}S_{i,i}-S_{i,i}D_{i,i}=0$$

اما $DS-SD$ قطری است پس $DS-SD=0$. در نهایت با توجه به $S^T=-S$ داریم

$$AA^T-A^TA=$$
$$(D+S)(D^T+S^T)-(D^T+S^T)(D+S)$$
$$SD+DS^T-DS-S^TD$$
$$=2(SD-DS)=0$$

مجله ریاضی شریف از هرگونه همکاری در تمامی زمینه‌ها از جمله تهیه یا معرفی مطالب علمی و توصیفی و همچنین همکاری در زمینه‌های اجرایی مجله، از جانب دانشجویان و اساتید استقبال به عمل می‌آورد. لازم به ذکر است که اکثر همکاران فعلی این مجله به صورت کاملا داوطلبانه همکاری می‌کنند و اساس کار این نشریه بر مبنای همکاری داوطلبانه اهالی دانشکده ریاضی شکل گرفته است.

تماس با ما:

mathematicsjournal@gmail.com

http://hamband.math.sharif.ir/journal