فرهنگ و اندیشهٔ ریاضی

سال ۲۷، شمارهٔ ۲، پاییز ۱۳۸۷

(تاریخ انتشار: زمستان ۱۳۸۷)

شمارهٔ پیاپی: ۴۱

فهرست مطالب

فرهنگ و اندیشهٔ ریاضی هر سال در دو شماره (بهار و پاییز) منتشر و به اعضای حقیقی، حقوقی و مشترکین انجمن ریاضی ایران ارسال می‌شود.

علاقه‌مندان به عضویت حقیقی و دانشگاه‌ها، مؤسسات و کتابخانه‌ها که تمایل به عضویت حقوقی یا اشتراک سالانه دارند می‌توانند با دبیرخانهٔ انجمن ریاضی ایران تماس حاصل نمایند.

شماره‌های قبلی این مجله با هماهنگی دبیرخانه قابل فروش می‌باشند.

روی جلد: راجر پنروز

فرهنگ و اندیشهٔ ریاضی
شماره ۴۱ (پاییز ۱۳۸۷) صص. ۱ تا ۷

فریب و ریاکاری با داده‌ها: کلاهبرداری در علوم*

دیوید هند

ترجمهٔ حمید پزشک

دو سال قبل جامعهٔ علمی از شنیدن خبر متهم شدن «وو ساک هانگ[۱]» محقق کلونینگ و سلول‌های بنیادی به جعل نتایج کارش، بسیار حیرت زده شد. نه فقط او، که زیگموند فروید[۲]، نیوتن[۳]، و برخی از برندگان جایزهٔ نوبل هم متهم به تقلب در داده‌ها شده‌اند. دیوید هند[۴] با بررسی مسألهٔ تقلب در علوم درمی‌یابد که این کار چندان هم نادر نیست.

وقتی «وو ساک هانگ» دانشمند بیومدیکال کره‌ای ادعا کرد توانسته با موفقیت شبیه‌سازی گاوها و یک سگ را به انجام برساند، توجه مطبوعات و مراکز علمی را به خود جلب نمود. این توجه با پیشرفت‌های مهم دیگری در مورد تحقیق سلول‌های بنیادی در آزمایشگاه او ادامه یافت. (Http://en.wikipedia.org/wiki/Hwang-Woo-Suk #-note-16)

ولی به‌زودی این توجهات تغییر جهت دادند زیرا همکار آمریکایی او جرالد شتن[۵] به دلیل آن که از روند اهدای تخمک انسانی راضی نبود از ادامهٔ تحقیقات با هانگ امتناع ورزید. بررسی‌ها نشان داد که او یا شرکای نزدیکش به زنان برای اهدای تخمکشان پول داده‌اند. با بالا گرفتن موضوعات اخلاقی، مسائل سرعت گرفتند. در ماه مارس ۲۰۰۶ او از شغلش در دانشگاه بین‌المللی سئول اخراج شد و در ماه مِی آن سال متهم به کلاهبرداری گشت. اعضای گروه تحقیق دریافتند هانگ، بسیاری از داده‌ها در مورد سلول‌های بنیادی را خود به وجود آورده و متهم شد که ۳ میلیون دلار از ۴۰ میلیون دلار پول تحقیق خود را به طور نادرست برای چیزهایی مثل خرید ماشین برای

---

*) این نوشته ترجمه‌ای از مقالهٔ زیر است:

Hand, D; (2007) Deception and Dishonesty with Data: Fraud in Science. Significance, issue 1, Vol. 4, pp 22-25

1) Woo Suk Hwang 2) Sigmund Freud 3) Newton 4) David Hand 5) Gerald Schatten

همسرش و هدیه برای سیاستمداران مصرف کرده است. او اکنون منتظر دادگاه است و احتمالاً به زندان محکوم خواهد شد. گاهی مردم یک تصور ترجیحاً ایده‌آل از دانشمندان دارند. آنان اغلب به عنوان مشاهدهٔ گرانی بی‌احساس از طبیعت دیده می‌شوند که بدون آشفته‌شدن از انگیزه‌های متناقض، اسرار را بیرون می‌آورند و قادرند آنچه را که موجب ناراحتی ما می‌شود، هدایت کنند. به هرحال حقیقت، همان طور که هر کسی در دنیای علوم کار کند می‌داند، این است که دانشمندان هم انسان هستند و مثل بقیه در معرض تأثیرات مشابه قرار دارند. محیط‌های علمی هم محیط‌هایی انسانی هستند و دانشمندان مثل هر کس دیگری، رؤسایی با درخواست‌های غیرممکن دارند، دارای دستیارانی هستند که آنچه را از آنها خواسته شده انجام نمی‌دهند و با آن شرایط داورانی متخصص هم دربارهٔ آنها قضاوت می‌کنند. مخصوصاً دنیای علم همانقدر به شدت رقابتی است که دنیای تجارت و بازرگانی.

رقابت، از راه‌های متعددی آشکار می‌شود. پیشرفت شغلی به مقالات موفق فرد بستگی دارد و لااقل در زمینه‌های دانشگاهی به جوایز نقدی. موفقیت، شناخت گروه متخصص، شاید قدرت، گاهی حتی وعدهٔ ثروتمندان را که با یک امتیاز به موقع پشتیبانی می‌شود، ایجاب می‌کند. حتی می‌تواند نوع اشتهاری را تعیین نماید که از انتساب یک اکتشاف یا پیشرفت به یک شخص حاصل می‌شود. تعجبی ندارد که بعضی محققین با توسل به کلاهبرداری می‌خواهند مطمئن شوند که موفق هستند. برای هانگ جوایز و فشارها شدیدتر از خیلی‌های دیگر بود. به او قبل از سقوطش، عنوان رسمی اولین دانشمند بی‌نظیر کره‌ای داده شد و یک سری تمبر به افتخار او چاپ شده بود.

در مقیاس کوچک‌تر، دیگران نیز تسلیم وسوسه شده‌اند. مارتینسون و دیگران در مقاله‌ای در Nature دریافتند که از ۳۲۴۷ دانشمند بررسی شده به طور ناشناس، ۳/۰٪ به داشتن داده‌های دروغین در کار تحقیقی خود اعتراف کرداند. ٦٪ هم اعتراف کردند که اگر داده‌ای با تحقیق قبلی‌شان ضدیت داشته، آن را ارائه نکرده‌اند. دانشمندان در اواسط دوران کاری خود، احتمالاً بیش از جوانترها در این رفتارهای نادرست زیاده‌روی کرده‌اند که شاید به دلیل وجود فشار بیشتر روی آنان در این مرحله از دوران کار باشد.

هرچند در بعضی از کلاهبرداری‌ها مثل کلاهبرداری بانکی، مرتکبین از همان ابتدا غیرصادق هستند و عمداً در معرض فریب خوردن قرار می‌گیرند، من فکر می‌کنم بیشتر دانشمندان متقلب به‌تدریج خود را اسیر کلاهبرداری می‌بینند. برای فهمیدن دلیل این کار، به داده‌هایی که دانشمندان جمع می‌کنند فکر کنید.

داده‌ها، حتی قبل از ارائهٔ به یک آماردان برای تحلیل، اغلب خود نتیجهٔ یک فرایند انتخاب هستند. محققی که در زمان اندازه‌گیری، یک ابزار یا نمونه را کنار می‌گذارد ممکن است احساس کند که اندازه‌گیری مربوطه باید حذف شود. اگر آنها متوجه شوند که یکی از کنترل کننده‌ها از مقدار خواسته شده متفاوت می‌باشد، ممکن است احساس کنند که نتیجه نباید در تحلیل وارد شود. در واقع اگر متوجه شوند که مقدار پاسخ متغیر از یکی از اندازه‌گیری‌های تصادفی کاملاً با دیگران متفاوت

است، حتی اگر هیچ توضیحی برای آن نداشته باشند، ممکن است شک کنند و آن را از آنالیز حذف نمایند. هیچ قانونی در مورد تعیین حد و حدود این کار وجود ندارد. به ناچار مسائل معنوی وارد جریان می‌شوند و با تمرین آزادی بیشتر در انتخاب مقدارهایی که فرد می‌خواهد قبول کند، می‌تواند هر نتیجه‌ای که بخواهد به‌دست آورد و از هر فرضیه‌ای که بخواهد پشتیبانی کند.

رابرت میلیکان، برندهٔ جایزهٔ نوبل فیزیک سال ۱۹۲۳، به سبب اندازه‌گیری بار الکترون در قطرهٔ روغن شهرت دارد. آزمایش او یک پیروزی بزرگ است و نشان می‌دهد که الکتریسیته یک جریان نیست بلکه به صورت واحدهای شبه ذره‌ای جداگانه در حرکت است. بار الکتریکی می‌تواند فقط در مضرب‌های عدد کامل واحد پایه‌اش وجود داشته باشد و شکستن بارهای الکتریکی امکان ندارد. ولی با وجود این، موضوع خیلی بحث برانگیز بود. میلیکان (که در اواسط دوران کاری‌اش بود و چندان مشهور نبود) نیاز داشت توجه دیگران را به این موضوع جلب کند. پس به نفع میلیکان بود که داده‌اش، از نظریهٔ او پشتیبانی کند و بسیاری از مشاهدات منتشر شدهٔ او نیز این کار را می‌کردند. به هرحال، همهٔ اندازه‌گیری‌های ثبت شدهٔ وی این طور نبودند. مثلاً در مقاله‌ای در Physical Review گزارش داد که «آزمایش او نه در مورد یک گروه انتخاب شده از قطره‌ها بلکه حاصل کار بر روی همهٔ قطره‌ها در ۶۰ روز متوالی است.» ولی او حقیقت را نمی‌گفت. بررسی دفترچه‌های یادداشت اصلی او توسط رابرت هولتون در سال ۱۹۷۸ نشان داد بسیاری از مشاهدات ثبت شدهٔ وی در بین آن آنالیزها وجود ندارند. میلیکان کار بر روی ۵۸ قطرهٔ کوچک را گزارش نمود ولی در واقع ۱۷۵ قطره ثبت کرد. یک نظریهٔ خوشبینانه این خواهد بود که او ۱۱۷ نتیجه را حذف کرده است،

(http://www.sigmaxi.org/meeting/archive/forum.2000.millikan.shtml)

نه به دلیل آن که این نتایج بر ضد فرضیه‌اش بودند بلکه به دلیل پایین بودن کیفیت داده‌ها این کار را کرد، با این که در دفترچه یادداشت خودش نوشته بود «این تقریباً کاملاً درست است» ، «این نتیجهٔ قشنگ را منتشر کن»، «خیلی پائین، چیزی غلط است»، «عدم توافق، درست در نمی‌آید.» این که چقدر جدی چنین چیزهایی مورد توجه قرار می‌گیرند، اغلب به این بستگی دارد که بعداً تحقیق از نظریه پشتیبانی کند. بدبختانه برای دانشمندی که می‌خواهد داده‌ها را به نظریه مرتبط کند علم یک فرایند خودکنترل است. نظریات غیرصحیح می‌توانند مدتی دوام بیاورند ولی بالاخره در مقابل مدارک مخالف خود، تسلیم می‌گردند.

سیر رویدادها برای «پیلت‌داون مَن[1]» نیز به همین صورت بود. در سال ۱۹۱۲ در ساسِکس استخوان فک و جمجمه‌ای توسط «چارلز داوسن[2]» به دست آمد که ادعا می‌کرد (و در آغاز مورد قبول هم واقع شد) که فسیل حلقهٔ گمشدهٔ کمیاب بین انسان و میمون است. ولی بعد کنار گذاشته شد چون ثابت شد جعلی است و حتی قبل از ۱۹۵۳ کاملاً فراموش گردید. جمجمه فقط ۶۰۰ سال عمر داشت و استخوان فک متعلق به اورانگوتان بود. هنوز هم کاملاً معلوم نشده که آیا داوسون فسیل را اول دفن و سپس بیرون آورده است یا نه. همچنین در سال ۱۹۲۰ «مورد قورباغهٔ قابله[3]»

1) Piltdown Man 2) Charles Dawson 3) Case of the Midwife Toad

می‌خواست از نظریهٔ لامارکی یعنی توارث خصوصیات اکتسابی بر ضد داروینیسم پشتیبانی کند. ولی نوارهایی که ادعا می‌شد نتیجهٔ کلیدی می‌باشند، تقلبی از آب درآمدند. «کامرر» که آزمایش را انجام داد خودکشی کرد. بنابراین کلاهبرداری علمی یک پدیدهٔ جدید نیست. در واقع به قدر کافی باعث نگرانی چارلز باب‌بیج[1]، یکی از مؤسسین انجمن سلطنتی آمار بوده که اقدام به ایجاد یک طبقه‌بندی کند: «انواع متعددی از فریب وجود دارند که در علوم به کار رفته‌اند ولی تعداد کمی شناخته شده‌اند مگر آنها که جدیدتر هستند... این ها ممکن است تحت عنوان کلک زدن، جعل کردن، آرایش کردن و پختن طبقه‌بندی شوند.» او کلک زدن را به عنوان «قصدِ لو رفتن در آخر کار» توصیف کرد تا کسانی را که دنبال آن می‌روند مسخره کرده باشد، ولی جعل کردن را قصدِ دوام یافتن تا مدتی طولانی معنی کرد. آرایش کردن به معنای «کوتاه کردن مقداری از مشاهداتی که از حد متوسط، افراط کرده‌اند و چسباندن آنها به مشاهداتی که زیادی مختصرند» می‌باشد؛ و بالاخره پختن را به عنوان «انتخاب داده‌ای که از بحث فرد پشتیبانی کند» مشخص کرد، اتهامی که علیه میلیکان وارد شده است از این نوع می‌باشد.

البته همهٔ دانشمندانی که داده‌ها را با انتخاب تغییر شکل می‌دهند، از کارشان آگاه نیستند. آنان ممکن است از پدیدهٔ آماریی مثل اریبی انتخاب بی‌اطلاع باشند و شاید همهٔ خدمات و دوره‌های آموزشی کمکی در مورد آمار باید شامل توضیحاتی در مورد میزان جدی بودن آن باشد. یا شاید درک نکنند چه اتفاقی در حال شکل‌گیری است. مثل موردی با کشف قلابی رابطه بین خیال پردازی و واکسن سرخک، اوریون و سرخجه با آنالیز ابتدایی براساس آزمایش بر روی یک گروه بسیار انتخابی کودکان. چنین مواردی نتیجهٔ کلاهبرداری نخواهد بود بلکه صرفاً نتیجهٔ جهل آمار پایه‌ای هستند.

اگر انتخاب داده‌ها می‌تواند از یک طرف بدون جهت گیری باشد و از طرف دیگر عمداً دستکاری شده باشد، پس قادر است خود اعداد را تغییر دهد. دستکاری صریح یعنی وقتی که فرد می‌گوید چه کاری انجام داده، کاملاً قابل قبول است. بنابراین دیگران می‌توانند تصمیم بگیرند که آیا به نظر آنها نتایج، تغییر شکل داده شده‌اند یا نه. ولی دستکاری پنهانی یعنی وقتی فرد تغییر را پنهان می‌کند، چیز کاملاً متفاوتی است. هر چه ابزارهای نرم‌افزاری پیشرفت می‌کنند، این کار آسان‌تر می‌شود. وسوسهٔ تصحیح سطحی فقط چند مورد از مشاهدات برای این که P مقدار را بامعنا کنند تا مقاله منتشر شود، باید گاهی بسیار قوی باشد.

مشکل این است که علوم در خلأ ساخته نشده‌اند. داده‌ها فقط می‌توانند در زمینهٔ یک نظریه تفسیر شوند. بنابراین طبیعتاً باید تمایلی برای حذف کردن مقدارهایی که با نظریه جور نیستند وجود داشته باشد. میلیکان ممکن است دلایل قابل توجیهی برای انتخاب مشاهدات خاصی داشته ولی تصور کردن دیگران در حال انتخاب یا اصلاح داده‌ها با حداقل بی‌گناهی آسان نیست.
اعتماد به نفس کافی یا جهل می‌تواند به آسانی به این نظریه منتهی شود که من می‌دانم نظریه‌ام

1) Charles Babbage

درست است و من در درست کردن داده‌ها، برحق می‌باشم تا اشتباهات را برطرف کنم و کاری کنم نظریه‌ام قطعی‌تر به نظر آید. میلیکان فقط یک نمونه است. بررسی دقیق دفترچه یادداشت‌های حتی بزرگترین دانشمندان گذشته ثابت می‌کند که خیلی از آنان این کار را کرده‌اند. نیوتن، مندل، پاستور، گالیله و کپلر همه، ابرهای تردید در بالای سر خود داشته‌اند. شواهد قطعی نشان می‌دهند که زیگموند فروید در ادعاهایش به گمراهی رفته است.

البته، فرد می‌تواند از حد فقط انتخاب و اصلاح هم فراتر رود. در نهایت اگر داده‌ایی که با زحمت جمع‌آوری شده نظریهٔ فرد را تأیید نکند، با عوض شدن، می‌تواند اصلاح شود. در سال ۲۰۰۵، ادارهٔ بهداشت عمومی آمریکا[1] دریافت که استیون لیدن[2] با تحریف نمونه‌های DNA و ساختن ترکیبات دروغین برای آزمایشات انجام شده در آزمایشگاهش به گمراهی افتاده تا از یافته‌های ناقص ادعایی خود در فرایند بازسازی DNA پشتیبانی کند.

(http://ori.dhhs.gov/miscoduct/cases/Leadon.shtml.)

حتی یک گام بلندتر هم می‌توان برداشت. وقتی کسی آنقدر در این راه لغزنده پایین رفته که داده‌ها را خود انتخاب و مقدارها را برای پشتیبانی از نظریه‌اش اصلاح یا تحریف کند، کافی است یک گام کوچک بردارد تا بفهمد چیزها آسان‌تر می‌بودند اگر از همان ابتدا داده‌ها را خود ابداع می‌کرد. در زمینه‌ای دیگر، مورد «سنگ جدول پیاده‌رو» برای توصیف سرشماری کننده‌هایی به کار می‌رود که روی سنگ جدول پیاده‌رو می‌نشینند و شمارش می‌کنند.

این، همان مورد «جان دارسی[3]» است. او یک متخصص داروسازی برای کمک به معالجهٔ سکتهٔ قلبی در هاروارد بود، که عده‌ای از همکارانش به او مشکوک شدند. سرپرست آزمایشگاه از او در مورد داده‌هایش پرس و جو کرد تا نتایج را بررسی کند ولی دارسی به جای ارائهٔ داده‌ها، شروع به اندازه‌گیری‌های جدید کرد، او همه اندازه‌گیری‌ها را در یک روز انجام داد ولی در روزهای مختلفی ثبت نمود. بررسی بعدی نشان داد داده‌های مقالات دیگر حتی زمانی که او هنوز دانشجوی دوره کارشناسی بود نیز دستکاری و یا اختراع شده بودند.

جان هندریک شون[4] نمونهٔ جدید دیگری در زمینهٔ فیزیک می‌باشد. او نیز مانند دارسی به نظر می‌آمد یک محقق خلاق درسال ۲۰۰۱ باشد که هر ۸ روز یک مقاله مشترک می‌نوشت. ولی دیگر محققین در نتایج کار او متوجه کارهای خلاف آشکاری شدند.

مخصوصاً که این نتایج بیش از حد درست به نظر می‌رسیدند. بررسی‌ها آشکار کردند که دو آزمایش متفاوت، مقدارهای مشابهی به دست آورده بودند و بررسی دقیق‌تر موارد دیگری را نشان داد که همان داده‌ها در آزمایشات دیگری گزارش شده بودند. معلوم شد که سری داده‌های دیگر به طور تجربی جمع‌آوری نشده ولی توسط کامپیوتر و با فرایندهای ریاضی حاصل آمده است.

کلاهبرداری علمی عواقبی دارد. در یک سطح دیگر، در کار یک کلاهبردار پیشرفتی دیده می‌شود

1) US Public Health Service 2) Steven Leadon 3) John Darsee 4) Jan Hendrik Schön

که شاید استحقاق آن را نداشته باشد. ولی گرفتاری‌های بسیار جدی‌تری وجود دارند. کلاهبرداری در تحقیقات پزشکی باعث بروز ضایعات مستقیم در مردمی می‌شود که از داروهای نامناسب استفاده می‌کنند. یک مقاله لیدون که در سال ۱۹۹۷ منتشر شده ۲۲۷ بار مورد ارجاع داشت تا این که بالأخره ملغی شد. این موضوع نشان می‌دهد یک کلاهبرداری اگر کشف نشود چه اثر مخربی می‌تواند داشته باشد.

همچنین تأثیر بر علوم و تصور خود علوم نیز وجود دارد. تبلیغات در مورد دانشمندان کلاهبردار می‌تواند از خود کلاهبرداری اثر مخرب بیشتری داشته باشد (که به همین دلیل، اغلب پنهان می‌شود ــ همان‌طور که یک بانکدار زمانی به من گفت: «بانک من هیچ نوع کلاهبرداری ندارد.» که همین، آن را منحصربه‌فرد می‌سازد). در جنگی که علم در مواجهه با ادعاهای بی‌معنای شبه‌علمی دارد آخرین چیزی که نیاز دارد موارد عمومی شده‌ای است که به این اشاره می‌کند که قهرمانان آن قابل اعتماد نیستند. از طرف دیگر، ممکن است وقتی یک کلاهبرداری کشف می‌شود، فریادهای ناشی از تعجب و شلوغ‌کردن زیاد، این عقیده را تقویت می‌کند که این امر خیلی بعید است.

ولی چقدر بعید است؟ کلاهبرداری چقدر در علوم شایع است؟ ماهیت کلاهبرداری جواب دادن به این پرسش را بسیار مشکل می‌سازد. حداقل یک تعریف دقیق از معنای کلاهبرداری را می‌طلبد که با چند مدل‌سازی آماری پیچیده پشتیبانی شود به طوری که با دیدن نوک کوه یخ، اندازهٔ آن را تخمین بزنند. چند مورد اخیر با این توافق خاتمه یافته‌اند که به شرط عدم پذیرش مسؤولیت تحقیقات علمی دیگر، هیچ عملی علیه فرد دانشمند صورت نگیرد. (در تمام مواردی که من از آنها آگاه شده‌ام، فرد دانشمند مذکر بوده است.) این ثابت می‌کند که خیلی از موارد اصلاً گزارش نمی‌شوند.

مقاله‌ای در Nature تعداد دانشمندانی که مرتکب جعل، تحریف یا سرقت ادبی می‌شوند را بین ۱٪ تا ۲٪ تخمین زده است.

سؤال این است که چگونه تقلب علمی را آشکار کنیم. یک ویژگی مشترک بین بسیاری از موارد کشف شده این است که اول یک خلاف کوچک کشف می‌شود و بررسی دقیق‌تر، خلاف بزرگتری را آشکار می‌کند که هرچه آزمایشات ادامه می‌یابند عمیق‌تر می‌شوند تا این که بالأخره تخلفات فاصله‌دار بزرگی معلوم می‌شوند. بعضی وقت‌ها تاریخچه‌ای از ابزارهای آماری دقیق وجود دارد که برای آشکارسازی کلاهبرداری به کار می‌روند. مانند مورد جان هندریک شون (و نیز نیوتن و مندل) که در آنها داده‌ها بیش از آن خوب بودند که درست باشند. آنها بیش از حد با فرضیه‌های تجربی منطبق بودند. گاهی داده‌ها به قدر کافی تصادفی نیستند (انسان‌ها در ساختن اعداد تصادفی خیلی خوب کار نمی‌کنند.) و آزمایش‌های آماری این را آشکار می‌کنند. مثلاً توزیع بنفورد[1]، توزیعی را به دست می‌دهد که برای اولین رقم در مجموعه‌های مشاهدات طبیعی تحت شرایط خاص انتظار خواهیم داشت و خروج از این توزیع می‌تواند ثابت کند اعداد تصنعی‌اند.

1) Benford

انسان چه باید بکند اگر به کلاهبرداری مشکوک شد؟ این به وضوح مشکل ساز است اگر یک محقق پایین‌تر به سرپرست خود مشکوک شود. برای یک کلاهبردار رده بالا غریب نیست که اگر متهم شد آن را پاک کند. ووساک هانگ، دیگران را در تهیهٔ داده‌های غلط مقصر می‌دانست. برعکس آن نیز دشوار است. مقالات توضیح می‌دهند داده‌های تحریف شدهٔ جان دارسی، با همکاری افراد پایین‌تر از او نوشته شده‌اند. آیا آنها باید برای عدم تشخیص کار خلاف جریمه شوند؟ تا چه حد می‌توان در کار دیگران دقت کرد؟ بالاخره جایی باید اعتماد وجود داشته باشد.

افراد و ادارات مختلفی برای کنترل کلاهبرداری علمی به کار گرفته شده‌اند. US Department of-Health and Human Services اداره‌ای در مورد درستی تحقیقات دارد، که در سال ۱۹۹۲ به عنوان ترکیبی از دو سازمان تأسیس شد و در سال ۱۹۹۳ از سازمان‌های مالی مستقل شد. در بریتانیا Committee on Publication Ethics سازمانی برای ناشرین مجلات علمی است که مسائل اخلاقی در تحقیقات را بررسی می‌کند.

بدبختانه با وجود چنین سازمان‌هایی از آنجا که علم یک حرفهٔ انسانی است مهم نیست با چه دقتی مراقبت و کنترل بشود، کلاهبرداری هرگز کاملاً ناپدید نمی‌شود. مانند فقر، کلاهبرداری همیشه با ما خواهد بود.

مراجع

[1] Martinson, B. C., Anderson, M.S. and de Vries, R. (2005) Scientists behaving badly. *Nature*, 435, 737-738.

[2] Koestler, A. (1973) The Case of the Midwife Toad. New York: Random House.

مترجم: حمید پزشک
دانشکده ریاضی، آمار و علوم کامپیوتر، قطب بیومتمتیک پردیس علوم، دانشگاه تهران
pezeshk@khayam.ut.ac.ir

.

فرهنگ و اندیشهٔ ریاضی
شماره ۴۱ (پاییز ۱۳۸۷) صص. ۹ تا ۲۶

اندازه‌های عدم اطمینان

سید محمود طاهری

چکیده

در این مقاله، نخست اندازه‌های عدم اطمینان، و به‌ویژه ردهٔ بزرگی از آن‌ها موسوم به اندازه‌های فازی، مورد بررسی قرار می‌گیرند. سپس دو نوع مهم این اندازه‌ها، یعنی اندازه‌های باور (اعتقاد) و اندازه‌های موجه‌نمایی، و حالت خاص آن‌ها یعنی اندازه‌های احتمال، مطالعه شده و ارتباط این اندازه‌ها با مفهومی به نام تابع تخصیص پایه توضیح داده می‌شود. به علاوه دو ردهٔ مهم دیگر از این اندازه‌ها، به نام‌های اندازه‌های امکان و اندازه‌های لزوم (حتمیت)، بررسی می‌گردند. مطالب ارائه شده، با مثال‌های عددی و کاربردی تشریح شده‌اند.

واژه‌های کلیدی: اندازه‌های عدم اطمینان، اندازه‌های فازی، اندازه‌های غیر جمعی، اندازه‌های باور، اندازه‌های امکان، اندازه‌های احتمال، تابع تخصیص پایه

۱. مقدمه و تاریخچه

از دیرباز تنها رهیافت تکامل یافتهٔ ریاضی برای اقدام در شرایط عدم اطمینان، نظریهٔ احتمال بوده است. اما از چند دههٔ گذشته به این سو، نظریه‌های گوناگون دیگری برای بررسی متغیرها و سیستم‌هایی که اطلاع نسبت به آن‌ها کافی و دقیق نیست، ارائه شده‌اند. بسیاری از این نظریه‌ها در یک چارچوب کلی موسوم به اندازه‌های عدم اطمینان[1] قرار می‌گیرند. این اندازه‌ها در واقع توابع مجموعه‌ای با ویژگی‌های خاص هستند. توسیع و تحدید این ویژگی‌ها منجر به انواع اندازه‌های عدم اطمینان می‌شود. گرچه اندازه‌های عدم اطمینان انواع گوناگونی دارند، اما ردهٔ بزرگ و مهمی از آن‌ها که اندازه‌های فازی[2] نامیده شده‌اند، بیشتر مورد توجه قرار گرفته‌اند.

1) Uncertainty Measures 2) Fuzzy Measures

همان طور که می‌دانیم، مفهوم اندازه یکی از کلیدی‌ترین مفاهیم در ریاضیات است. اندازه‌های متداول در ریاضیات اندازه‌های جمعی (اندازه‌های با ویژگی جمع‌پذیری) هستند. این ویژگی در عین حال که در بسیاری از موارد مفید و راهگشاست، ولی به هر حال، یک محدودیت تلقی می‌شود. اندازه‌های فازی در جهت رفع این محدودیت معرفی شده‌اند. این اندازه‌ها تعمیمی از مفهوم متداول اندازه‌اند، به این معنی که در آن‌ها شرط جمع پذیری توسط شرط ضعیف‌تر یکنوایی جایگزین می‌شود. چون غیرجمعی بودن، ویژگی اصلی اندازه‌های فازی است، این اندازه‌ها گاهی اندازه‌های غیرجمعی[1] نیز نامیده می‌شوند.

از لحاظ تاریخی، اندازه‌های فازی نخستین بار توسط سوگینو در سال ۱۹۷۲ [۳۷] (بر گرفته از رساله دکترای وی [۳۸]) مطرح شد. البته پیش از آن تلاش‌هایی در جهت ارائهٔ روش‌هایی جدید برای توصیف و تحلیل عدم اطمینان انجام گرفته بود، که از جمله می‌توان به مفاهیم احتمال‌های بالایی و پایینی[2] توسط دمپستر [۹] و مطالعهٔ احتمال در چارچوب اندازه‌های باور (اعتقاد)[3] توسط شفر [۳۵] اشاره کرد. (برای مروری بر مبانی این دو نظریه به [۱] مراجعه کنید).

پس از معرفی اندازه‌های فازی توسط سوگینو، این مبحث به یکی از مباحث جدی و گسترده در نظریه‌های اندازه، نظریه‌های عدم اطمینان و نظریه‌های اطلاع، و کاربرد آن‌ها تبدیل شده است.

دوبوا و پراد [۱۱] و مورفوشی و سوگینو [۲۷ و ۲۸] مروری جامع بر مفاهیم و ویژگی‌های مرتبط با اندازه‌های فازی ارائه داده‌اند. در این زمینه، کتاب‌های دننبرگ [۱۰] و وانگ و کلیر [۴۲] دربارهٔ اندازه‌های فازی، و به طور کلی اندازه‌های غیر جمعی، حائز اهمیت است.

باندمر و نتر [۴]، گودمن و گوین [۱۵] و گرابیش و همکاران [۱۷] اندازه‌های فازی را با تأکید بر جنبه‌های کاربردی مطالعه نموده‌اند. لیگینلال و او [۲۴] ضمن مرور برخی کاربردهای اندازه‌های عدم اطمینان، کاربردی از این اندازه‌ها را در فرایند تصمیم‌گیری مطالعه نموده‌اند. نیز ناروکاوا و تورا [۲۹] کاربرد اندازه‌های فازی را در نظریهٔ بازی‌ها بررسی کرده‌اند. فرسون و گینزبرگ [۱۴]، هن کیند و هاریسون [۱۸]، کلیر [۱۹] و کلیر و فولگر [۲۱] به بررسی انواع اندازه‌های عدم اطمینان و تفاوت‌های بین آن‌ها پرداخته‌اند. انتقال از یک چارچوب عدم اطمینان به چارچوبی دیگر، توسط افرادی مانند دوبوا و پراد [۱۳] و کلیر [۲۰] بررسی شده است. برونویچ و کارکیشچنکف [۶] ساختار اندازه‌های فازی تولید شده توسط احتمال‌های بالایی و پایینی را بررسی کرده‌اند. لوکس و اعرابی [۲۶] یک اندازهٔ فازی مقدار، مبتنی بر تعمیمی از قضیه دمپستر ــ شفر، ارائه کرده‌اند. لی و ژو [۲۳] قواعد و شیوه‌های استنتاج را بر اساس برخی از اندازه‌های فازی مطالعه نموده‌اند. تفاوت‌های بین اندازه‌های احتمال و امکان به وسیلهٔ گودمن و گوین [۱۶]، رالسکو و رالسکو [۳۲] و فیتل و هارتر [۴۰]، و مبحث احتمال‌های نادقیق توسط بوادری و همکاران [۷]، والی [۴۱] و زاده [۴۴] مطالعه شده است. اندازه‌های عدم اطمینان با رویکرد مبتنی بر مجموعه‌های تصادفی توسط گوین [۳۰]

1) Non-additive Measures 2) Upper and Lower Probabilities 3) Belief Measures

بررسی شده است. نیز این اندازه‌ها، با تأکید بر نظریهٔ اعتبار[1] توسط لیو [۲۵] مطالعه شده است.

موضوع انتگرال‌های غیرجمعی و به‌ویژه انتگرال‌های فازی که مرتبط و متناسب با اندازه‌های غیرجمعی تعریف می‌شوند، توسط محققانی مانند بن‌ونوتی و مسیار [۵]، مورفوشی و سوگینو [۲۸]، رالسکو و آدامز [۳۳]، اشمیدلر [۳۴] و سوگینو [۳۹] مورد مطالعه قرار گرفته است. خاطر نشان می‌شود که انتگرال‌های غیرجمعی، در یک تقسیم‌بندی کلی، به چهار شیوه شامل انتگرال شوکه [۸]، انتگرال سی‌پوس [۳۶]، انتگرال سوگینو و انتگرال T ـ همنرم رده‌بندی می‌شوند. یک اثر کامل در این باره کتاب دننبرگ [۱۰] است. نیز اثر پپ [۳۱] که در آن ردهٔ وسیعی از توابع مجموعه‌ای مقدار و انتگرال چنین توابعی را مطالعه نموده است، شایان توجه است.

مقالهٔ حاضر بدین ترتیب تدوین شده است: در بخش دوم تعریف اندازه‌های فازی ارائه می‌شود و ضمن بیان چند مثال، تفاوت آن با اندازه‌های احتمال و مجموعه‌های فازی تشریح می‌گردد. در بخش سوم، دو نوع خاص از اندازه‌های فازی، موسوم به اندازه‌های باور و اندازه‌های موجه‌نمایی (مقبولیت)[2]، بررسی می‌گردند. در بخش چهارم مفهوم تابع تخصیص (واگذاری) پایه[3]، که یک مفهوم کلیدی در مبحث اندازه‌های عدم اطمینان است، معرفی می‌شود. بخش پنجم به اندازه‌های احتمال و بخش ششم به اندازه‌های امکان[4] و اندازه‌های لزوم[5] که حالت‌های خاص و مهمی از اندازه‌های فازی هستند، اختصاص دارد. مراجع اصلی این مقاله [۱۷] و [۲۱] و [۲۸] هستند.

۲. اندازه‌های فازی

در این بخش و در بخش‌های آینده فرض می‌کنیم که X یک مجموعهٔ مرجع و $\mathcal{P}(X)$ مجموعهٔ توانی آن و $\mathcal{B}(X)$ (به کوتاهی: $\mathcal{B}$) یک میدان سیگمایی از زیرمجموعه‌های X است.

تعریف ۱.

تابع مجموعه‌ای $g:\mathcal{B}\longrightarrow[0,1]$ را یک اندازهٔ فازی بر $(X,\mathcal{B})$ گوییم اگر در سه اصل زیر صدق کند:

[g_1 :] (شرایط مرزی)

$$g(\varnothing)=0 \quad , \quad g(X)=1$$

[g_2 :] (یکنوایی)

$$A\subseteq B\Longrightarrow g(A)\leq g(B) \quad , \quad \forall\, A,B\in\mathcal{B}$$

1) Credibility Theory 2) Plausibility Measures 3) Basic Assignment Function
4) Possibility Measures 5) Necessity Measures

[$g_۳$:] (پیوستگی) برای هر دنبالهٔ یکنوا از زیرمجموعه‌های X، یعنی برای هر دنباله $\{A_i \in \mathcal{B} \ , \ i \in N\}$ که

$$A_۱ \subseteq A_۲ \subseteq \cdots \quad \text{یا} \quad A_۱ \supseteq A_۲ \supseteq \cdots$$

داشته باشیم

$$\lim_{i \longrightarrow \infty} g(A_i) = g(\lim_{i \longrightarrow \infty} A_i)$$

در تعریف بالا هر عضو $\mathcal{B}$ یک پیشامد[۱] نام دارد و مجموعهٔ مرجع X فضایی است شامل پیشامدهایی که با دسته‌ای اطلاعات نادقیق و نامطمئن دربارهٔ عناصر X مرتبط‌اند، و $g(A)$ درجهٔ اطمینان نسبت به رخ دادن پیشامد A است.

اصل اول بیان می‌کند که، صرف نظر از اطلاعات[۲] و شواهدی[۳] که در اختیار داریم، هر عنصر مورد بررسی مطمئناً در X قرار دارد و مطمئناً در مجموعهٔ تهی قرار ندارد.

اصل دوم بیانگر این احساس شهودی است که اگر پیشامد A پیشامد B را نتیجه دهد $(A \subseteq B)$، آن گاه حداقل همان اطمینان را که به رخ دادن A داریم، به رخ دادن B داریم. به سخن دیگر، اصل دوم نتیجه می‌دهد که اطلاعات و شواهد مربوط به عضویت یک عنصر در مجموعهٔ B باید دست‌کم به اندازهٔ شواهد عضویت آن عنصر به هر زیرمجموعه از B، مانند A، باشد.

اصل سوم، که فقط برای حالتی کاربرد دارد که X نامتناهی است، لزوم پیوسته بودن تابع g را بیان می‌کند. به سخن دیگر این اصل نوعی سازگاری را تضمین می‌کند: محاسبه $g(A)$ $(A = \lim_{i \longrightarrow \infty} A_i)$ چه به‌صورت حد $g(A_i)$ ها و چه با به‌کارگیری g برای حد A_i ها به یک نتیجه منجر شود.

تذکر ۱. برخی محققان تعریف اندازهٔ فازی را با شرایط ضعیف‌تری ارائه نموده‌اند. از جمله در برخی متون در تعریف اندازهٔ فازی، شرط $g(X) = ۱$ ذکر نمی‌شود و اندازهٔ فازی با چنین شرطی، اندازهٔ فازی نرمال شده[۴] نامیده می‌شود. برای توضیح بیشتر به [۱۷] رجوع کنید.

مثال ۱. می‌خواهیم نوع بیماری یک مریض را تشخیص دهیم. برای سادگی، فرض کنید می‌خواهیم بدانیم که مریض فعلی واجد کدام یک از بیماری‌های برونشیت، سرماخوردگی یا سینه پهلوست. آزمایش‌های پزشکی و عوارض بیماری، اطلاعات و شواهدی در اختیار می‌گذارند که به تشخیص بیماری کمک می‌کند. این اطلاعات آن اندازه دقیق و مطمئن و کامل نیستند که بتوان نوع بیماری را به‌طور قطعی تعیین کرد. اما می‌توان بر پایهٔ این اطلاعات درجهٔ اطمینان نسبت به وجود هر بیماری را مشخص نمود. مثلاً ممکن است، طبق اطلاعات و شواهد موجود، به اندازه ۷۰% مطمئن شویم که مریض عضو مجموعه انسان‌هایی است که برونشیت دارند، و ۵۰% مطمئن شویم

1) Event 2) Information 3) Evidence 4) Normalized Fuzzy Measure

که وی عضو مجموعه افرادی است که سرماخوردگی دارند و ۲۰٪ مطمئن شویم که وی سینه پهلو دارد. این مقادیر تشکیل یک اندازه فازی را می‌دهند که توصیف کنندهٔ عدم اطمینان نسبت به نوع بیماری فرد مورد نظر است.

مثال ۲. دادگاهی را در نظر بگیرید که برای یک محاکمه جنایی دربارهٔ بی‌گناهی یا مجرمیت تعدادی متهم تشکیل شده است. دادگاه برای قضاوت دربارهٔ این افراد متوسل به اطلاعات، شواهد و گواه‌هایی می‌شود. دادگاه نمی‌تواند به‌طور قطع مجرمیت یا بی‌گناهی هر متهم را تعیین کند. اما می‌تواند، طبق شواهد و گواه‌های موجود، تعیین کند که با چه اندازه اطمینانی هر متهم عضو مجموعهٔ مجرمین است. این نوع عدم اطمینان را می‌توان در قالب اندازه‌های فازی توصیف و تبیین کرد.

نکتهٔ ۱. از تعریف ۱ و مثال‌های بالا آشکار است که نوع عدم اطمینانی که با اندازه‌های فازی تبیین می‌شود با نوع عدم اطمینانی که توسط مجموعه‌های فازی بیان می‌شود، متفاوت است. در حالت اخیر، با یک مجموعهٔ فازی روبرو هستیم یعنی مجموعه‌ای با کران‌های نادقیق (مثلاً مجموعه افراد بلند قد، مجموعه اعداد طبیعی کوچک، مجموعه اعداد حدوداً ۱۰، ...). آنگاه به هر شیء درجه‌ای نسبت می‌دهیم که بیانگر درجهٔ عضویت آن شیء در مجموعهٔ فازی مورد نظر است. اما در مبحث اندازه‌های فازی با مجموعه‌هایی دقیق (غیر فازی) روبرو هستیم. در این جا عدم اطمینان به این موضوع باز می‌گردد که، طبق اطلاعات و شواهد موجود، چه اندازه مطمئن هستیم که یک فرد (شیء) خاص عضو هر مجموعه است. برای نمونه در مثال تشخیص پزشکی، مجموعه افراد دارای برونشیت یک مجموعه مشخص و غیرفازی فرض می‌شود. حال برپایهٔ اطلاعات و شواهد موجود، به اندازه ۷۰٪ اطمینان داریم که مریض فعلی عضو مجموعه افراد واجد برونشیت است. هم چنین در مثال محاکمهٔ جنائی، مجموعه افراد مجرم و مجموعه افراد بی‌گناه دو مجموعهٔ دقیق و غیرفازی هستند. در این جا عدم اطمینان به این باز می‌گردد که تا چه اندازه مطمئن هستیم که یک فرد خاص عضو مجموعه افراد مجرم است (یا این که تا چه اندازه مطمئن هستیم که هر مجموعه خاص از افراد، شامل فرد یا افراد مجرم است).

پس در حالت مجموعه‌های فازی، تابع عضویت تابعی از مجموعه مرجع X به بازه $[\text{۰}, \text{۱}]$ است. در حالی که در مبحث اندازه‌های فازی، هر اندازه فازی تابعی از $\mathcal{B}$ (یک میدان سیگمایی از X) به بازه $[\text{۰}, \text{۱}]$ است.

نکتهٔ ۲. تفاوت اندازه‌های فازی و اندازه‌های احتمال شایان توجه است. در واقع اصول موضوعه‌ای که در تعریف ۱ بیان شده است، ضعیف‌تر از اصول موضوعهٔ احتمال است. بدین ترتیب می‌توان گفت که اندازه‌های احتمال حالت خاصی از اندازه‌های فازی هستند. تفاوت اساسی به اصل جمع‌پذیری احتمال باز می‌گردد. اصلی که بیان می‌دارد که برای هر مجموعه از پیشامدهای جدا از هم A_i

$$P(\bigcup_{i\in I} A_i) = \sum_{i\in I} P(A_i)$$

که در آن I یک مجموعهٔ شمارای اندیس گذار است. از همین اصل نتیجه می‌شود که $P(A^C) = ۱ - P(A)$. به سخن دیگر اندازهٔ احتمال متمم هر پیشامد را می‌توان برحسب اندازهٔ احتمال آن پیشامد به‌دست آورد. اما در اندازه‌های فازی با دانستن $g(A)$ (درجه اطمینان به رخ دادن پیشامد A) مقدار $g(A^C)$ (درجه اطمینان به رخ ندادن A) را نمی‌توان دقیقاً تعیین کرد.

نتیجهٔ ۱. از اصل یکنوایی ($g_۲$) نتیجه می‌شود که برای هر $A, B \in \mathcal{B}$،

$$g(A \cap B) \leq \min[g(A)\ ,\ g(B)] \qquad (۱.\text{الف})$$

$$g(A \cup B) \geq \max[g(A)\ ,\ g(B)] \qquad (۱.\text{ب})$$

۳. اندازه‌های باور و اندازه‌های موجه‌نمایی

تعریف ۲. تابع مجموعه‌ای $Bel : \mathcal{B} \longrightarrow [۰, ۱]$ را یک اندازهٔ باور بر $(X, \mathcal{B})$ گوییم اگر در اصول $g_۱$ تا $g_۳$ صدق کرده و به علاوه برای هر $n \in N$ و هر رده از زیرمجموعه‌های $A_۱, A_۲, \ldots, A_n$ از $\mathcal{B}$،

$$Bel(A_۱ \cup A_۲ \cup \cdots \cup A_n) \geq \sum_i Bel(A_i) - \sum_{i<j} Bel(A_i \cap A_j) + \cdots + (-۱)^{n+۱} Bel(A_۱ \cap \cdots \cap A_n) \qquad (۲)$$

برای نمونه در حالت $n = ۲$ رابطهٔ فوق به این صورت در می‌آید

$$Bel(A_۱ \cup A_۲) \geq Bel(A_۱) + Bel(A_۲) - Bel(A_۱ \cap A_۲) \qquad (۳)$$

مقدار $Bel(A)$ درجهٔ باوری است که بر پایهٔ اطلاعات و گواه‌ها و شواهد موجود، به رخ دادن پیشامد A داریم (تعبیر دقیق‌تر را به زودی خواهیم گفت). اصل ممیزهٔ اندازه‌های باور، رابطهٔ ۲، بیان می‌دارد که درجهٔ باور مرتبط با اجتماع چند پیشامد جدا از هم، از مجموع درجات باور تک تک پیشامدها کمتر نباشد. پس این اصل حالت ضعیف‌تری از اصل جمع پذیری احتمال است. بنابراین اندازه‌های احتمال حالت خاصی از اندازه‌های باور هستند: در اندازه‌های احتمال نامساوی ۲ تبدیل به تساوی می‌شود. (برای تعمیمی از اندازه‌های باور [۲۲] را ببینید).

نتیجهٔ ۲. اگر در رابطهٔ ۲ قرار دهیم $n = ۲$ و $A_۱ = A$ و $A_۲ = A^C$، خواهیم داشت

$$Bel(A) + Bel(A^C) \leq ۱$$

به هر اندازهٔ باور، یک اندازهٔ موسوم به اندازهٔ موجه‌نمایی مرتبط است که برای هر مجموعهٔ A از $\mathcal{B}$ به‌صورت زیر تعریف می‌شود:

$$Pl(A) = ۱ - Bel(A^C)$$

و به‌طور مشابه

$$Bel(A) = ۱ - Pl(A^C)$$

بنابراین اندازه‌های باور و موجه‌نمایی، دوگان یکدیگرند. اندازه‌های موجه‌نمایی را می‌توان مستقلاً و به‌صورت زیر نیز تعریف نمود.

تعریف ۳. تابع مجموعه‌ای $Pl : \mathcal{B} \longrightarrow [۰, ۱]$ را یک اندازهٔ موجه‌نمایی بر $(X, \mathcal{B})$ گوییم اگر در اصول $g_۱$ تا $g_۳$ صدق کرده و به‌علاوه برای هر $n \in N$ و هر رده از زیرمجموعه‌های $A_۱, A_۲, \ldots, A_n$ از $\mathcal{B}$،

$$Pl(A_۱ \cap A_۲ \cap \cdots \cap A_n) \leq \sum_i Pl(A_i) - \sum_{i<j} Pl(A_i \cup A_j) + \cdots + (-۱)^{n+۱} Pl(A_۱ \cup \cdots \cup A_n) \quad (۴)$$

برای نمونه در حالت $n = ۲$ رابطهٔ فوق به این صورت در می‌آید

$$Pl(A_۱ \cap A_۲) \leq Pl(A_۱) + Pl(A_۲) - Pl(A_۱ \cup A_۲) \quad (۵)$$

که برای مقایسه با رابطهٔ ۳ می‌توان آن را چنین نوشت

$$Pl(A_۱ \cup A_۲) \leq Pl(A_۱) + Pl(A_۲) - Pl(A_۱ \cap A_۲)$$

اصل ممیزهٔ اندازه‌های موجه‌نمایی (رابطهٔ ۴) حالت ضعیف‌تری از اصل جمع‌پذیری احتمال است. بنابراین اندازه‌های احتمال حالت خاصی از اندازه‌های موجه‌نمایی هستند: در اندازه‌های احتمال نامساوی ۴ تبدیل به تساوی می‌شود.

نتیجهٔ ۳. اگر در رابطهٔ ۴ قرار دهیم $n = ۲$ و $A_۱ = A$ و $A_۲ = A^C$، خواهیم داشت

$$Pl(A) + Pl(A^C) \geq ۱$$

۴. تابع تخصیص پایه

تعریف ۴. تابع مجموعه‌ای $m : \mathcal{B} \longrightarrow [۰, ۱]$ را یک تابع تخصیص پایه بر $(X, \mathcal{B})$ گوییم اگر $m(\varnothing) = ۰$ و

$$\sum_{A \in \mathcal{B}} m(A) = ۱ \quad (٦)$$

مقدار $m(A)$ درجه‌ای از اطمینان است که به رخ دادن دقیقاً پیشامد A (و نه هیچ زیرمجموعهٔ مشخص از آن) داریم.

نکتهٔ ۳. معادلهٔ ٦ شبیه به معادلهٔ مربوط به توابع جرم احتمال، $\sum_{x \in X} P(x) = ۱$، است. اما تفاوتی اساسی بین توابع جرم احتمال و توابع تخصیص پایه وجود دارد. توابع جرم احتمال بر X تعریف

می‌شوند و توابعی عددی هستند، در حالی که تخصیص‌های پایه بر عناصر $\mathcal{B}$ تعریف می‌شوند و توابعی مجموعه‌ای هستند. خاطر نشان می‌شود که در بعضی منابع از عبارت تابع تخصیص احتمال پایه استفاده شده است. ما، به‌ویژه برای تمایز با توابع احتمال، از عبارت تابع تخصیص پایه استفاده می‌کنیم. نیز در برخی از متون به جای تابع تخصیص پایه، عبارت تابع جرم[1] را به کار می‌برند.

نکتهٔ ۴. گاهی بهتر است شرط $m(\varnothing) = 0$ در تعریف ۴ را حذف نمود. این کار برای حالت‌هایی توصیه می‌شود که مجموعهٔ اطلاعات در مورد پدیدهٔ مورد بررسی کامل نیست و با نوعی نقصان اطلاعات روبرو هستیم. (نیز رجوع کنید به تذکر ۱).

تعریف ۵. هر مجموعهٔ $A \in \mathcal{B}$ که برای آن $m(A) > 0$، یک عنصر کانونی m (*Focal Element*) نامیده می‌شود.

عناصر کانونی، زیرمجموعه‌هایی از مجموعهٔ مرجع X هستند که اطلاعات و شواهد و قرائن و گواه‌های موجود، بر آن‌ها تمرکز دارند. به‌ویژه هنگامی که X متناهی باشد، می‌توانیم تابع m را با فهرست عناصر کانونی و اندازه‌های آن‌ها مشخص کنیم.

نکتهٔ ۵. با توجه به تعریف تابع تخصیص پایه، در می‌یابیم که اولاً لزومی ندارد که $m(X) = 1$، و دوم آن که تابع تخصیص پایه یک تابع یکنوا نیست یعنی لزومی ندارد که اگر $A \subset B$ آنگاه $m(A) \leq m(B)$ و سوم آن که رابطه‌ای خاص بین $m(A)$ و $m(A^C)$ وجود ندارد.

گرچه تخصیص‌های پایه، اندازه‌های فازی نیستند اما نکتهٔ اساسی و مهم این است که هر تابع تخصیص پایه m بر $(X, \mathcal{B})$، به‌طور یکتا یک اندازهٔ باور و یک اندازهٔ موجه‌نمایی بر $(X, \mathcal{B})$ به صورت زیر تعیین می‌کند

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{۷}$$

$$Pl(A) = \sum_{B \cap A \neq \varnothing} m(B) \tag{۸}$$

بنابراین اگر در یک مسأله، مقادیر تابع تخصیص پایه (یا صرفاً عناصر کانونی و مقادیر تابع تخصیص پایه به‌ازای آن‌ها) را بدانیم، می‌توانیم درجه‌های باور و موجه‌نمایی هر پیشامد را محاسبه کنیم. گفتنی است که می‌توان جهتی مخالف با جهت بالا در تعریف توابع باور و موجه‌نمایی را در پیش گرفت. بدین ترتیب که نخست تابع تخصیص پایه را معرفی نمود و سپس توابع باور و موجه‌نمایی را به عنوان توابعی که در روابط ۷ و ۸ صدق می‌کنند، تعریف کرد.

1) Mass Function

نتیجهٔ ۴. از روابط ۷ و ۸ نتیجه می‌شود که برای هر پیشامد A،

$$m(A) \leq Bel(A) \leq Pl(A)$$

با روابط ۷ و ۸ می‌توان براساس تابع تخصیص پایه، اندازه‌های باور و موجه‌نمایی را به‌دست آورد. برای حالت عکس، یعنی هنگامی که یک اندازهٔ باور یا یک اندازهٔ موجه‌نمایی داده شده باشند، تخصیص پایهٔ متناظر طبق قضیهٔ زیر به‌دست می‌آید.

قضیهٔ ۱. (قضیهٔ وارون موبیوس) اگر یک اندازهٔ باور $Bel(.)$ یا یک اندازهٔ موجه‌نمایی $Pl(.)$ داده شده باشد، تخصیص‌های پایهٔ متناظر این گونه به‌دست می‌آیند

$$m(A) = \sum_{B\subseteq A} (-1)^{|A-B|} Bel(B) \tag{۹}$$

$$m(A) = \sum_{B\subseteq A} (-1)^{|A-B|} (1 - Pl(B^C)) \tag{۱۰}$$

که در آن $|A-B|$ عدد اصلی مجموعهٔ $A-B$ است.

اثبات. رابطهٔ ۹ را ثابت می‌کنیم. از رابطهٔ ۷ داریم

$$\begin{aligned} \sum_{B\subseteq A} (-1)^{|A-B|} Bel(B) &= \sum_{B\subseteq A} (-1)^{|A-B|} \sum_{C\subseteq B} m(C) \\ &= \sum_{C\subseteq A} m(C) + \sum_{B\subset A} (-1)^{|A-B|} \sum_{C\subseteq B} m(C) \\ &= m(A) + \sum_{C\subset A} m(C) + \sum_{B\subset A} (-1)^{|A-B|} \sum_{C\subseteq B} m(C) \end{aligned}$$

اما

$$\sum_{C\subset A} m(C) + \sum_{B\subset A} (-1)^{|A-B|} \sum_{C\subseteq B} m(C) = 0$$

زیرا همهٔ جمله‌های عبارت $\sum_{C\subset A} m(C)$ با علامت منفی در عبارت بعد ظاهر می‌شود. به‌علاوه جملات باقیمانده در عبارت دوم نیز به‌طور دوتایی با علامت‌های منفی و مثبت ظاهر می‌شوند. دقت کنید که $m(\varnothing) = 0$. بدین ترتیب رابطهٔ ۹ ثابت می‌شود. رابطهٔ ۱۰ نیز با جایگذاری از رابطهٔ $Bel(B) = 1 - Pl(B^C)$ به‌دست می‌آید.

مقایسهٔ اندازه باور، اندازهٔ موجه‌نمایی و مقدار تخصیص پایه

با توجه به تعریف تابع تخصیص پایه و روابط ۷ و ۸، بر پایهٔ مجموعه‌ای از اطلاعات و شواهد دربارهٔ مجموعهٔ مرجع X و زیرمجموعه‌های آن، می‌توان گفت که

$m(A)$ درجهٔ اطمینانی است که به رخ دادن دقیقاً پیشامد A (و نه هیچ زیرمجموعهٔ مشخصی از آن) داریم.

$Bel(A)$ درجهٔ اطمینانی است که به رخ دادن پیشامد A و یا هر زیرمجموعه‌ای از آن داریم.

$Pl(A)$ درجهٔ اطمینانی است که برای رخ دادن پیشامد A و یا هر زیرمجموعه از آن و یا هر پیشامد دیگری که با A اشتراکی دارد، قائل هستیم.

تعبیرهای فوق را می‌توان از دیدگاه نظریهٔ اطلاع نیز بیان کرد. فرض کنید براساس اطلاعات موجود می‌خواهیم دربارهٔ عضویت یک شیء (فرد) مشخص از X در مجموعهٔ A نظر دهیم. در این حالت

$m(A)$ بیانگر درجهٔ گواهی اطلاعات موجود بر عضویت آن شیء در دقیقاً مجموعهٔ A است.

$Bel(A)$ بیانگر درجهٔ گواهی اطلاعات موجود بر عضویت آن شیء در مجموعهٔ A و یا هر زیرمجموعه از A است.

$Pl(A)$ بیانگر درجهٔ گواهی اطلاعات موجود بر عضویت آن شیء در مجموعه A و یا هر زیرمجموعه از A و یا هر مجموعه دیگری که با A اشتراک دارد، می‌باشد.

مثال ۳. یک فرد بیمار به پزشک مراجعه می‌کند. پزشک بعد از معاینه و آزمایش‌های لازم نوع بیماری را یکی از سه نوع برونشیت A، سرماخوردگی B، و سینه پهلو C تشخیص می‌دهد و براساس نتایج به‌دست آمده، تخصیص پایه m مندرج در جدول ۱ را ارائه می‌دهد. براساس تخصیص پایه m، و با استفاده از رابطه‌های ۹ و ۱۰ اندازه‌های باور و موجه‌نمایی متناظر را نیز می‌توانیم به‌دست آوریم.

عناصر کانونی	m	Bel	Pl
A	۰/۱۵	۰/۱۵	۰/۸۵
B	۰	۰	۰/۵۵
C	۰/۱۰	۰/۱۰	۰/۷۵
$A \cup B$	۰/۱۰	۰/۲۵	۰/۹۰
$A \cup C$	۰/۲۰	۰/۴۵	۱
$B \cup C$	۰/۰۵	۰/۱۵	۰/۸۵
$A \cup B \cup C$	۰/۴۰	۱	۱

جدول ۱. تخصیص پایه و اندازه‌های باور و موجه‌نمایی نوع بیماری در مثال ۳

توضیح. پزشک پس از معاینات و آزمایش‌های لازم، ملاحظه می‌کند که ۱۵٪ از علائم و آثار بیماری، مختص بیماری برونشیت، A، است و هیچ علامتی که مختص بیماری سرماخوردگی، B، باشد وجود ندارد و به علاوه ۱۰٪ از علائم بیماری نیز مختص بیماری سینه پهلو، C، می‌باشد. هم‌چنین ۱۰٪ از آثار بیماری، از علائم مشترک بین بیماری‌های A و B است (و نه هر کدام از آن‌ها به‌طور خاص) و ... هم چنین تا آخر، یعنی ۴۰٪ از علائم بیمار، از علائم مشترک بین هر سه بیماری است. به سخن دیگر ۴۰٪ از اطلاعات موجود صرفاً گواهی بر این دارند که بیمار در مجموعهٔ بیماران مبتلا به برونشیت یا سرماخوردگی و یا سینه پهلو قرار دارد و از این ۴۰٪ اطلاعات هیچ نتیجهٔ خاص دیگری عاید نمی‌شود. هم‌چنین مثلاً مقدار ۴۵٪ برای $Bel(A \cup C)$ این گونه تفسیر می‌شود که درجهٔ باور ما به این که بیمار مبتلا به برونشیت و یا مبتلا به سینه پهلوست برابر ۴۵٪ است. و هم‌چنین مقدار ۸۵٪ برای $Pl(B \cup C)$ به این معنی است که ۸۵٪ علائم بیماری، متعلق به بیماری سرماخوردگی و یا سینه پهلو و یا از علائم مشترک این دو بیماری و یا از علائمی است که مشترک با علائم سرماخوردگی و یا سینه پهلو و بیماری دیگر است.

ملاحظه می‌کنید که نوع اطلاعات و دانش حاصله و نوع اندازه‌های عدم اطمینان که در این جا مطرح است و نوع تخصیص‌دهی مقادیر به مجموعه‌های مختلف، اساساً با آنچه در چارچوب‌های احتمالی بیان می‌شود (و عمدتاً مبتنی بر فراوانی نسبی است) تفاوت بنیادی دارد.

۵. اندازه‌های احتمال

اگر اصل ممیزه اندازه‌های باور یعنی رابطهٔ ۲ با اصل قوی‌تر زیر جایگزین شود، حالت خاصی از اندازه‌های باور که همان اندازه‌های احتمال است به‌دست می‌آید

$$Bel(A \cup B) = Bel(A) + Bel(B) \quad , \quad A \cap B = \varnothing \qquad (۱۱)$$

قضیهٔ زیر اساسی‌ترین ویژگی اندازه‌های احتمال را به‌عنوان حالت خاصی از اندازه‌های باور بیان می‌دارد [۲۱].

قضیهٔ ۲. یک اندازه باور $Bel(.)$ بر $(X, \mathcal{P}(X))$، که در آن $\mathcal{P}(X)$ متناهی است، یک اندازهٔ احتمال است اگر و تنها اگر تخصیص پایهٔ متناظر آن، برای مجموعه‌های تک عضوی برابر اندازهٔ باور آن‌ها باشد، یعنی

$$m(\{x\}) = Bel(\{x\})$$

و برای هر مجموعهٔ غیر تک عضوی A، $m(A) = ۰$.

از قضیهٔ بالا نتیجه می‌شود که اندازه‌های احتمال بر مجموعه‌های مرجع متناهی، به وسیلهٔ یک تابع به‌صورت $P : X \longrightarrow [۰, ۱]$ تعریف می‌شوند به قسمی که $P(x) = m(\{x\})$. همان‌طور که می‌دانیم در نوشتارهای آماری این تابع، تابع جرم احتمال (*Probability Mass Function*) نام دارد.

نتیجهٔ ۵. اگر تخصیص پایه فقط بر مجموعه‌های تک عضوی متمرکز شود، داریم $\sum_{x\in X} m(\{x\}) = 1$. در این صورت طرف‌های راست معادله‌های ۷ و ۸ برابر می‌شوند و خواهیم داشت

$$Bel(A) = Pl(A) = \sum_{x\in A} m(\{x\}) = \sum_{x\in A} P(x) = P(A), \quad A \in \mathcal{P}(X)$$

این بدان معنی است که اندازه‌های دوگان باور و موجه‌نمایی تحت اصل ۱۱ به اندازه‌های احتمال تبدیل می‌شوند.

نکتهٔ ٦. گفته شد که اندازه‌های احتمال هنگامی دقیقاً تعیین می‌شوند که عناصر کانونی m، مجموعه‌های تک عضوی باشند. در حالت کلی ممکن است چنین نباشد. در این حالت گرچه نمی‌توان احتمال هر پیشامد را دقیقاً مشخص کرد اما می‌توان، تحت شرایطی، آن را به‌طور نادقیق و به‌صورت یک بازه بیان کرد. کران بالای این بازه اندازهٔ موجه‌نمایی و کران پایین آن اندازهٔ باور مربوط به آن پیشامد است.

توجه کنید که اگر m یک تابع تخصیص پایه بر $\mathcal{B}$ باشد، و $Bel(.)$ و $Pl(.)$ اندازه‌های باور و موجه‌نمایی متناظر با m باشند، آن گاه می‌توان مجموعه‌ای از اندازه‌های احتمال بر $\mathcal{B}$ در نظر گرفت که در رابطهٔ زیر صدق کنند

$$Bel(A) \le P(A) \le Pl(A) \quad , \quad A \in \mathcal{B}$$

این رابطه توضیح می‌دهد که چرا در برخی شرایط، از $Bel(A)$ و $Pl(A)$ به‌عنوان احتمال‌های پایینی و بالایی نام برده می‌شود. احتمال‌های پایینی و بالایی نخستین بار توسط دمپستر [٦] در سال ۱۹٦۷، و البته مستقل از مفهوم اندازه‌های باور و موجه‌نمایی، معرفی شد. برای مطالعهٔ این رویکرد به احتمال، مراجع [۹]، [۳۰] و [۴۱] را ببینید. خوانندهٔ علاقمند به رویکردهای مختلف به احتمال، در مقام یک اندازهٔ عدم اطمینان و تفسیرها و تعبیرهای گوناگون احتمال می‌تواند به مراجع [۲] و [۳] مراجعه کند.

٦. اندازه‌های امکان و لزوم

هنگامی که عناصر کانونی تو در تو[1] باشند آن گاه اندازه‌های باور و موجه نمایی مربوطه، هماهنگ[2] نامیده می‌شوند. این بدان معنی است که اطلاعات و شواهد مربوط به یک مسئله، با یکدیگر در تزاحم نیستند. در این حالت روابط خاصی برای اندازه‌های باور و موجه‌نمایی برقرار می‌شود که در قضیهٔ زیر ارائه شده است [۱۸].

قضیهٔ ۳. فرض کنید m یک تابع تخصیص پایه بر $(X, \mathcal{B})$ باشد. اگر عناصر کانونی m تو در تو باشند آنگاه برای هر $A, B \in \mathcal{B}$،

---

1) Nested 2) Consonant

$$Bel(A \cap B) = \min[Bel(A)\ ,\ Bel(B)] \qquad \text{(۱۲.الف)}$$
$$Pl(A \cup B) = \max[Pl(A)\ ,\ Pl(B)] \qquad \text{(۱۲.ب)}$$

در این حالت خاص، اندازه‌های باور و موجه‌نمایی به ترتیب اندازه‌های لزوم و امکان نامیده می‌شوند. این اندازه‌ها را می‌توان مستقیماً به‌عنوان دو نوع خاص از اندازه‌های فازی نیز تعریف کرد. (نیز رجوع کنید به [۹]).

تعریف ٦. تابع مجموعه‌ای $N : \mathcal{B} \longrightarrow [۰, ۱]$ را یک اندازه لزوم بر $(X, \mathcal{B})$ گوییم اگر در اصول $g_۱$ تا $g_۳$ صدق کند و به‌علاوه برای هر $A, B \in \mathcal{B}$،

$$N(A \cap B) = \min[N(A)\ ,\ N(B)] \qquad (۱۳)$$

ملاحظه می‌کنید که اندازه‌های لزوم از تحمیل تساوی در رابطهٔ (۱.الف) حاصل می‌شود. حالت حدی دیگر، از تحمیل تساوی در رابطه (۱.ب) به‌دست می‌آید.

تعریف ۷. تابع مجموعه‌ای $\Pi : \mathcal{B} \longrightarrow [۰, ۱]$ را یک اندازه امکان بر $(X, \mathcal{B})$ گوییم اگر در اصول $g_۱$ تا $g_۳$ صدق کند و به‌علاوه برای هر $A, B \in \mathcal{B}$،

$$\Pi(A \cup B) = \max[\Pi(A)\ ,\ \Pi(B)] \qquad (۱۴)$$

توضیح آن که، برپایهٔ اطلاعات و شواهد موجود، $N(A)$ درجهٔ حتمیت و لزوم رخ دادن پیشامد A است. هم‌چنین $\Pi(A)$ درجهٔ سازگاری اطلاعات و شواهد موجود با پیشامد A است. توجه دارید که برای هر $A \in B$،

$$\Pi(A) \geq N(A)$$

رابطهٔ اخیر با این درک شهودی منطبق است که یک پیشامد قبل از آن که لازم و حتمی شود، ممکن می‌شود.

واضح است که اندازه‌های لزوم و امکان دوگان یکدیگر هستند و به‌طور یکتا توسط رابطه‌های زیر از هم حاصل می‌شوند

$$\Pi(A) = ۱ - N(A^C) \qquad \text{(۱۵.الف)}$$
$$N(A) = ۱ - \Pi(A^C) \qquad \text{(۱۵.ب)}$$

به سخن دیگر، امکان یک پیشامد برابر است با یک منهای اندازهٔ حتمیت متمم آن پیشامد، و اندازهٔ حتمیت یک پیشامد برابر با یک منهای اندازه امکان متمم آن پیشامد است.

نتیجهٔ ٦. برای هر اندازهٔ لزوم و امکان روابط زیر برقرار هستند

$$\min[N(A)\ ,\ N(A^C)] = ۰ \qquad \text{(۱٦.الف)}$$
$$\max[\Pi(A)\ ,\ \Pi(A^C)] = ۱ \qquad \text{(۱٦.ب)}$$

اثبات. رابطهٔ ۱۶. الف براساس رابطهٔ ۱۳ و با توجه به اصل g_1 حاصل می‌شود. رابطهٔ ۱۶.ب نیز بر پایهٔ رابطهٔ ۱۴ و با توجه به اصل g_1 به‌دست می‌آید.

رابطهٔ ۱۶.الف بیان می‌کند که هیچ گاه، از دو پیشامد متناقض، هر دو اندازه‌های لزوم مثبت ندارند. رابطهٔ ۱۶.ب بیان می‌دارد که از بین دو پیشامد A و پیشامد متمم A، دست کم یکی کاملاً ممکن است رخ دهد.

برای هر اندازهٔ لزوم و اندازهٔ امکان دوگان آن، روابط زیر برقرار هستند

$$N(A) > 0 \implies \Pi(A) = 1 \qquad \text{(۱۷.الف)}$$

$$\Pi(A) < 1 \implies N(A) = 0 \qquad \text{(۱۷.ب)}$$

اثبات. اگر $N(A) > 0$ آنگاه از رابطهٔ ۱۶.الف نتیجه می‌شود که $N(A^C) = 0$. اکنون، با استفاده از رابطهٔ ۱۵.الف داریم $\Pi(A) = 1$. رابطهٔ ۱۷.ب به‌طور مشابه (نیز با برهان خلف و با استفاده از رابطهٔ ۱۷.الف) ثابت می‌شود.

بر پایهٔ آنچه گفته شد می‌توان رابطه‌های متناظر در اندازه‌های امکان و اندازه‌های لزوم را به‌صورتی که در جدول زیر درج شده است خلاصه کرد. برای مطالعهٔ بیشتر در مورد اندازه‌های امکان به [۱۲] و [۳۰] و [۴۳] مراجعه کنید.

روابط مربوط به اندازه‌های لزوم	روابط مربوط به اندازه‌های امکان
$N(A \cap B) = \min[N(A), N(B)]$	$\Pi(A \cup B) = \max[\Pi(A), \Pi(B)]$
$N(A) = 1 - \Pi(A^c)$	$\Pi(A) = 1 - N(A^c)$
$\min[N(A), N(A^c)] = 0$	$\max[\Pi(A), \Pi(A^c)] = 1$
$N(A) + N(A^c) \leq 1$	$\Pi(A) + \Pi(A^c) \geq 1$
$N(A) > 0 \Rightarrow \Pi(A) = 1$	$\Pi(A) < 1 \Rightarrow N(A) = 0$
$\Pi(A) \geq N(A)$	

منابع و مراجع

[۱] اعرابی، ب؛ شهبازی میرزا حسنلو، ط (۱۳۸۶)، مروری بر مبانی نظریه دمپستر شفر، اندیشه آماری، ش ۲۳، ص ۱۹—۳.

[۲] طاهری، س م (۱۳۸۱)، یگانگی و چندگانگی احتمال، نامه فرهنگستان علوم، ش ۱۹، ص ۱۲۵—۹۳.

[۳] گیلیز، د (۱۳۸۶)، نظریه‌های فلسفی احتمال، ترجمه: مشکانی، م ر، مؤسسهٔ انتشارات علمی دانشگاه صنعتی شریف.

[4] Bandemer, H., Näther, W. (1992); Fuzzy Data Analysis, Kluwer, Dordrecht.

[5] Benvenoti, P., Mesiar, R. (2000); Integrals with respect to a general fuzzy measure, In: Grabisch, M., et al. (Eds.); Fuzzy Measures and Integrals, Physica-Verlag, pp. 205 - 232.

[6] Bronevich, A.G., Karkishchenko, A.N. (2002); The structure of fuzzy measure families induced by upper and lower probabilities, In: Bertoluzza, C., et al. (Eds.), Statistical Modeling, Analysis and Management of Fuzzy Data, Physica-Verlag, pp. 160-172.

[7] Buadrit, C., Dubois, D., Fargier, H. (2004); Representation of incomplete probabilistic information, In: Lopez-Diaz, et al. (Eds.), Soft Methodology and Random Information Systems, Springer, pp. 149 - 156.

[8] Choquet, G. (1953); Theory of capacities, Ann. Inst. Fourier, 5: 131 - 295.

[9] Dempster, A.P. (1967); Upper and lower probabilities induced by multivalued mappings, Ann. Math. Stat, 38: 325 - 339.

[10] Dennenberg, D. (1997); Non-additive Measure and Integral, 2nd ed., Kluwer, Dordrecht.

[11] Dubois, D., Prade, H. (1987); Properties of measures of information in evidence and possibility theories, Fuzzy Sets and Systems, 24: 161-182.

[12] Dubois, D., Prade, H. (1988); Possibility Theory, Plenum Press, New York.

[13] Dubois, D., Prade, H., Sandri, S. (1993); On possibility/probability transformations, In: Lowen, R., Roubens, M. (Eds.), Fuzzy Logic: State of the Art, Kluwer, pp. 104 - 112.

[14] Ferson, S., Ginzburg, L.R. (1996); Different methods are needed to propagate ignorance and variability, Reliability Engineering and Systems Safety, 54: 133 - 144.

[15] Goodman, I.R., Nguyen, H.T. (1985); Uncertainty Models for Knowledge-Based Systems, North-Holland, Amsterdam.

[16] Goodman, I.R., Nguyen, H.T. (2002); Fuzziness and randomness, In: Bertoluzza, C., et al. (Eds.), Statistical Modeling, Analysis and Management of Fuzzy Data, Physica-Verlag, pp. 3-21.

[17] Grabisch, M., Nguyen, H.T., Walker, E.A. (1995); Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference, Kluwer, Dordrecht.

[18] Henkind, S.J., Harrison, M.C. (1988); An analysis for four uncertainty calculi, IEEE Trans. on Systems, Man, and Cybernetics, 18: 700 - 714.

[19] Klir, G.J. (1987); Where do we stand on measures of uncertainty, ambiguity, fuzziness, and the like?, Fuzzy Sets and Systems, 24: 141 - 160.

[20] Klir, G.J. (1990); A principle of uncertainty and information invariance, Intern. J. of General Systems, 17: 249 - 275.

[21] Klir, G.J., Folger, T.A. (1988); Fuzzy Sets, Uncertainty, and Information, Prentice-Hall, Englewood Cliffs, N. J.

[22] Kramosil, I. (2000); Towards generalized belief functions, In: Grabisch, M., et al. (Eds.); Fuzzy Measures and Integrals, Physica-Verlag, pp. 104 - 123.

[23] Lee, E.S., Zhu, Q. (1995); Fuzzy and Evidence Reasoning, Physica-Verlag, Heidelberg.

[24] Liginlal, D., Ow, T.T. (2006); Modeling attitude to risk in human decision processes: An application of fuzzy measures, Fuzzy Sets and Systems, 157: 3040 - 3054.

[25] Liu, B. (2007); Uncertainty Theory, Sec. Ed., Springer-Verlag, Heidelberg.

[26] Lucas, C., Araabi, B.N. (1999); Generalization of the Dempster-Shafer theory: a fuzzy-valued measure, IEEE Trans. Fuzzy Systems, 7 (3): 255 - 270.

[27] Murofushi, T., Sugeno, M. (1991); A thoery of fuzzy measures: representations, the Choquet integral, and null sets, J. Math. Anal. Appl., 159: 532 - 549.

[28] Murofushi, T., Sugeno, M. (2000); Fuzzy measures and fuzzy integrals, In: Grabisch, M., et al. (Eds.), Fuzzy Measures and Integrals, Physica-Verlag, pp. 3 - 41.

[29] Narukawa, Y., Torra, V. (2007); Fuzzy measures and integrals in evaluation of strategies, Fuzzy Sets and Systems, 177: 4686 - 4695.

[30] Nguyen, H.T. (2006); An Introduction to Random Sets, Chapman and Hall/CRC, Boca Raton.

[31] Pap, S. (1995); Null-Additive Set Functions, Kluwer, Dordrecht.

[32] Ralescu, A.L., Ralescu, D.A. (1980); Probability and fuzziness, Inform. Sci., 34: 85 - 92.

[33] Ralescu, D.A., Adams, G. (1980); The fuzzy integral, J. Math. Anal. Appl., 75: 562 - 570.

[34] Schmeidler, D. (1986); Integral representation without additivity, Proc. Amer. Math. Soc., 97: 255 - 261.

[35] Shafer, G. (1976); A Mathematical Theory of Evidence, Princeton Univ. Pub.

[36] Sipos, J. (1979); Non linear integrals, Math. Slovaka, 29: 257 - 270.

[37] Sugeno, M. (1972); Fuzzy measure and fuzzy integral, Trans. SICE, 8: 95-102.

[38] Sugeno, M. (1974); Theory of fuzzy integrals and its applications, Doctoral Thesis, Tokyo Institute of Technology.

[39] Sugeno, M. (1977); Fuzzy measures and fuzzy integrals - a survey, In: Gupta, M.M., et al. (Eds.), Fuzzy Automata and Decision Processes, North - Holland, pp. 89 - 102.

[40] Viertl, A., Hareter, D. (2004); Fuzzy information and stochastics, Iranian J. Fuzzy Systems, 1: 43 - 56.

[41] Walley, P. (1991); Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London.

[42] Wang, Z., Klir, G.J. (1992); Fuzzy Measure Theory, Plenum Press, New York.

[43] Yager R.R. (1980); Aspects of possibilistic uncertainty, Intern. J. of Man-Machine Studies, 12: 283 - 298.

[44] Zadeh, L.A. (2002); Toward a perception-based theory of probabilistic reasoning with imprecise probabilities, J. Stat. Plan. Inf., 105: 233 - 264.

سید محمود طاهری
دانشکده علوم ریاضی، دانشگاه صنعتی اصفهان
Taheri@cc.iut.ac.ir

فرهنگ و اندیشهٔ ریاضی
شماره ۴۱ (پاییز ۱۳۸۷) صص. ۲۷ تا ۳۸

آشنایی با برنامه‌ریزی روی مخروط‌های درجه دوم

مازیار صلاحی

چکیده

در این مقاله با توجه به کاربرد روزافزون مسائل برنامه‌ریزی روی مخروط‌ها از جمله مخروط‌های درجه دوم به معرفی این مسائل می‌پردازیم. ابتدا مفاهیم پایه‌ای مربوط به این زمینه را بیان می‌نماییم. سپس به معرفی دوگان یک مسألهٔ برنامه‌ریزی روی مخروط درجه دوم می‌پردازیم و چند مثال متفاوت که قابل بیان شدن به صورت برنامه‌ریزی روی مخروط‌های درجه دوم هستند، معرفی می‌کنیم. در ادامه به معرفی جبر جردن روی مخروط‌های درجه دوم می‌پردازیم و با استفاده از آن شرایط بهینگی را برای مسائل مذکور بیان و اثبات می‌کنیم؛ سرانجام تفاوت‌های این مسائل را با مسائل برنامه‌ریزی خطی بیان می‌کنیم.

۱. مقدمه

معرفی روش سیمپلکس در سال ۱۹۴۷ توسط جرج دانتزیک و کاربردهای فراوان آن در علوم مختلف توجه دانشمندان را به استفاده از این شاخه جلب نمود [۵]. علی‌رغم کارایی بالای روش سیمپلکس در حل مسائل برنامه‌ریزی خطی با ابعاد بزرگ، و استفاده از آن در حل برخی مسائل غیرخطی، این روش قابل تعمیم برای همه مسائل موجود نبود [۷]. روش‌های جریمه‌ای و مانعی[1] دستهٔ دیگری از الگوریتم‌ها هستند که با موفقیت در حل مسائل غیرخطی مورد استفاده قرار گرفته‌اند [۷] و در حال حاضر نیز از جمله روش‌های موفق در این زمینه می‌باشند [۹]. ردهٔ خاصی از مسائل

1) Penalty and Barrier Methods

غیرخطی، برنامه‌ریزی روی مخروط‌های درجه دوم می‌باشد که مسائل زیادی مربوط به زمینه‌های مختلف همچون کنترل، سرمایه‌گذاری، علوم مهندسی، پزشکی و غیره را می‌توان به صورت آن بیان کرد [۳ ، ۴]. مثال‌های خاصی از این نوع مورد مطالعه قرار گرفته‌اند، به عنوان نمونه، مثال ۴ در بخش ۳ حالت کلی مسألهٔ مشهور فرما ــ وبر می‌باشد [۱۰]. مقالهٔ لوبو مثال‌های متعددی از علوم مهندسی را در بر دارد که به صورت برنامه‌ریزی روی مخروط‌های درجه دوم بیان می‌شوند [۸]. با توجه به فراوانی مسائلی که به صورت برنامه‌ریزی روی مخروط‌های درجه دوم قابل بیان هستند. امروزه به جای حل این مسائل با استفاده از الگوریتم‌های کلی که در حل مسائل غیرخطی مورد استفاده قرار می‌گرفتند، الگوریتم‌های خاصی با استفاده از روش‌های نقطهٔ درونی معرفی شده‌اند که در اکثر نرم‌افزارهای تجاری و آموزشی مورد استفاده قرار می‌گیرند [۲، ۹].

در این مقاله ابتدا به معرفی مفاهیم پایه‌ای برنامه‌ریزی روی مخروط‌های درجه دوم می‌پردازیم، سپس دوگان آن را بیان و مثال‌های متعددی را معرفی می‌کنیم که در چارچوب این برنامه قابل بیان هستند. در بخش ۴ معرفی جبر جردن روی مخروط‌های درجه دوم با استفاده از شرایط بهینگی مسأله صورت می‌پذیرد. همچنین با ذکر مثال‌هایی به تفاوت‌های نطریهٔ دوگانی آنها با برنامه‌ریزی خطی می‌پردازیم.

۲. آشنایی با مخروط‌های درجه دوم و خواص آنها

تعریف ۱.۲. گوییم مجموعهٔ $C \subseteq \mathbb{R}^n$ یک مخروط است هرگاه برای هر $x \in C$ و $\alpha \geq 0$ داشته باشیم $\alpha x \in C$. همچنین گوییم مجموعهٔ C یک مخروط محدب است هرگاه هم مخروط و هم محدب باشد.

به عنوان مثال مجموعه‌های زیر محدب هستند [۳]:

- فضای بردارهای نامنفی n بعدی R^n_+.
- $Q_n = \{(x_1; \bar{x}) \in R^n | x_1 \geq ||\bar{x}||\}$ که $\bar{x} = (x_2, ..., x_n)^T \in R^{n-1}$ و $x_1 \in R$. Q_n به مخروط درجه دوم[1] (مخروط لورنتز[2] و یا مخروط بستنی [3]) معروف است.
- فضای تمام ماتریس‌های نیمه معین مثبت و متقارن که با S^n_+ نشان داده می‌شود.

تعریف ۲.۲. C^* را دوگان مخروط C گوییم هرگاه

$$C^* = \{y : \ x^T y \geq 0, \ \forall x \in C\}.$$

همچنین مخروط C را خود دوگان می‌نامیم اگر $C^* = C$.

به عنوان مثال مخروط‌های $S^n_{+1}, Q_n, \mathbb{R}^n_+$ همگی خود دوگان هستند. در اینجا ثابت می‌کنیم که مخروط درجه دوم خود دوگان است. برای سایر موارد خواننده می‌تواند به [۴] مراجعه کند.

1) Second Order Cone 2) Lorentz Cone 3) ice cream Cone

لم ۱.۲ مخروط Q_n خود دوگان است.

اثبات. فرض کنید $(x_1;\bar{x}) \in Q_n$ دلخواه باشد. نشان می‌دهیم که $(x_1;\bar{x}) \in Q_n^*$. اگر $x_1 = 0$ آنگاه حکم واضح است. فرض کنید $x_1 \neq 0$ و $(y_1;\bar{y}) \in Q_n$ داریم $x_1y_1 + \bar{x}^T\bar{y} \geq x_1y_1 - ||\bar{x}||||\bar{y}|| \geq 0$. در نتیجه $(x_1;\bar{x}) \in Q_n^*$. $Q_n \subseteq Q_n^*$. برای اثبات رابطهٔ عکس، فرض کنید که $(x_1;\bar{x}) \in Q_n^*$. همچنین واضح است که $y = (1, 0, 0, ..., 0) \in Q_n$. حال از تعریف Q_n^* داریم که $x^Ty \geq 0$. بنابراین $x_1 \geq 0$. برای ادامهٔ اثبات دو حالت را در نظر می‌گیریم:

- $\bar{x} = 0$، که در این حالت حکم واضح است.
- $\bar{x} \neq 0$. ابتدا نشان می‌دهیم که $x_1 > 0$. فرض کنید چنین نباشد، یعنی $x_1 \leq 0$. چون از قبل داریم $x_1 \geq 0$ پس $x_1 = 0$. از طرفی می‌دانیم که $z = (||\bar{x}||; -\bar{x}) \in Q_n$. حال چون $x \in Q_n^*$ و $z \in Q_n$، پس از تعریف Q_n^* داریم که $x^Tz \geq 0$. از طرفی $x_1 = 0$، پس $x^Tz = -||\bar{x}||^2$. بنابراین $||\bar{x}|| = 0$ که با فرض $\bar{x} \neq 0$ در تناقض است. حال باید نشان دهیم که $x \in Q_n$. فرض کنیم چنین نباشد یعنی $x_1 < ||\bar{x}||$. بردار $u = (x_1, \frac{-x_1}{||\bar{x}||}\bar{x})$ را در نظر بگیرید. به آسانی می‌توان نشان داد که $u \in Q_n$. چون $u \in Q_n$ و $x \in Q_n^*$ پس داریم $u^Tx \geq 0$. از طرفی $u^Tx = x_1(x_1 - ||\bar{x}||) < 0$ که با نامساوی قبلی در تناقض است. بنابراین فرض این که $x_1 < ||\bar{x}||$ نمی‌تواند درست باشد، پس $x \in Q_n$. این نتیجه می‌دهد که $Q_n^* \subseteq Q_n$.

از خواص مهم دیگری که یک مخروط در برنامه‌ریزی روی مخروط باید داشته باشد این است که مخروط نوک تیز باشد [۳، ۴].

تعریف ۳.۲ گوییم مخروط C نوک تیز[1] است هرگاه C شامل هیچ خطی نباشد یا به عبارتی دیگر $C \cap \{-C\} = \{0\}$.

با استفاده از این خاصیت می‌توان همانند برنامه‌ریزی خطی یک ترتیب را روی مخروط مورد نظر تعریف کرد، زیرا بسیاری از خواص برنامه‌ریزی خطی از ترتیب $\geq$ روی R_+^n نتیجه می‌شود. بنابراین هروقت صحبت از برنامه‌ریزی روی مخروط به میان می‌آید منظور مخروطی است که محدب، نوک تیز، بسته و دارای درون ناتهی می‌باشد. قضیهٔ زیر برخی خواص مخروط دوگان را بیان می‌کند.

قضیهٔ ۱.۲ فرض کنید $C \subseteq R^n$ یک مخروط باشد.

- اگر $intC \neq \emptyset$ آنگاه C^* نوک تیز است.
- اگر C محدب، بسته و نوک تیز باشد آنگاه $intC^* \neq \emptyset$.

1) Pointed Cone

• اگر C بسته و محدب باشد آنگاه $C^{**} = C$ که C^{**} دوگان مخروط C^* است.

یک نتیجهٔ واضح قضیهٔ فوق به صورت زیر است:

نتیجهٔ ۱.۲ زیرمجموعهٔ دلخواه C از R^n یک مخروط محدب بسته و نوک تیز با درون ناتهی است اگر و تنها اگر C^* هم چنین باشد.

۳. برنامه‌ریزی روی مخروط‌های درجه دوم

در این بخش برای سادگی فرض می‌کنیم که مسألهٔ مورد نظر فقط دارای یک مخروط درجه دوم است. نتایج کلی را در انتهای مقاله می‌آوریم. مسألهٔ برنامه‌ریزی روی مخروط درجه دوم می‌تواند به صورت زیر بیان می‌شود:

$$\begin{aligned} &\min\ c^T x \\ &Ax = b \\ &x \in Q_n, \end{aligned} \tag{۱}$$

که در آن $c \in R^n$، $b \in R^m$، $A \in R^{m\times n}$، $x \in R^n$ متغیر مسأله و Q_n نیز یک مخروط درجه دوم در R^n می‌باشند. ملاحظه می‌کنیم که تنها تفاوت این مسأله با برنامه‌ریزی خطی در قید آخر می‌باشد که به جای مخروط R^n_+ مخروط Q_n در نظر گرفته شده است. حال با پیروی از آنچه برای پیداکردن دوگان یک مسألهٔ برنامه‌ریزی خطی انجام می‌شود، به معرفی دوگان مسألهٔ (۱) می‌پردازیم. با استفاده از نظریهٔ دوگانی لاگرانژ، دوگان این مسأله به صورت زیر می‌باشد:

$$\begin{aligned} &\max\ \ b^T y \\ &A^T y + z = c \\ &z \in Q_n. \end{aligned} \tag{۲}$$

قبل از بیان شرایط بهینگی این مسأله، مجموعه و تابعی را معرفی می‌کنیم که به صورت یک مخروط درجه دوم با بعد مناسب قابل بیان هستند. سپس به ذکر چند مسألهٔ برنامه‌ریزی غیر خطی می‌پردازیم که به صورت برنامه‌ریزی روی مخروط درجه دوم بیان می‌شوند.

مثال ۱: مخروط

$$K = \{(x, \sigma_۱, \sigma_۲) \in R^n \times R \times R |\ \sigma_۱, \sigma_۲ \geq ۰, x^T x \leq \sigma_۱\sigma_۲\}$$

$$x_۱^۲ + \cdots + x_n^۲ \leq \sigma_۱\sigma_۲ = (\frac{\sigma_۱ + \sigma_۲}{۲})^۲ - (\frac{\sigma_۱ - \sigma_۲}{۲})^۲$$

که به وضوح معادل است با

$$K = \{(x, \sigma_1, \sigma_2) \mid \|(x; \frac{\sigma_1 - \sigma_2}{2})\| \leq \frac{\sigma_1 + \sigma_2}{2}\}$$

یک مخروط درجه دوم است.

مثال ۲: تابع کسری زیر را در نظر بگیرید [۳]:

$$g(x,s) = \begin{cases} \frac{x^T x}{s}, & s > 0 \\ 0, & s = 0, x = 0 \\ +\infty, & \text{در غیر این صورت} \end{cases}$$

حال با فرض این که $\frac{x^T x}{0}$ برابر ۰ یا $+\infty$ است بسته به این که $x = 0$ یا $x \neq 0$ و با توجه به تساوی $ts = \frac{(t+s)^2}{4} - \frac{(t-s)^2}{4}$ داریم:

$$\{\frac{x^T x}{s} \leq t, s \geq 0\} \leftrightarrow \{x^T x \leq ts, t \geq 0, s \geq 0\}$$

$$\leftrightarrow \{x^T x + \frac{(t-s)^2}{4} \leq \frac{(t+s)^2}{4}, t \geq 0, s \geq 0\} \leftrightarrow \|(x; \frac{t-s}{2})\| \leq \frac{t+s}{2}.$$

به وضوح نابرابری اخیر به صورت نابرابری مخروط درجه دوم است.

مثال ۳: برنامه‌ریزی اکیداً محدب زیر را در نظر بگیرید [۱]:

$$\begin{aligned} \min \quad & x^T Q x + a^T x + \beta \\ & Ax = b \\ & x \geq 0. \end{aligned}$$

تابع هدف این مسأله را می‌توان به صورت $\|\bar{u}\|^2 + \beta - \frac{1}{4} a^T Q^{-1} a$ نوشت که $\bar{u} = Q^{\frac{1}{2}} x + \frac{1}{2} Q^{-\frac{1}{2}} a$.
حال مسألهٔ زیر را که به صورت برنامه‌ریزی روی مخروط است در نظر بگیرید:

$$\begin{aligned} \min \quad & u_1 \\ & Q^{\frac{1}{2}} x - \bar{u} = \frac{1}{2} Q^{-\frac{1}{2}} a \\ & Ax = b \\ & x \geq 0, (u_1; \bar{u}) \in Q_{n+1}. \end{aligned}$$

لازم به ذکر است که جواب بهینهٔ این مسأله با جواب بهینه مسألهٔ اصلی یکی است و تنها اختلاف آنها در مقدار ثابت $\beta - \frac{1}{4} a^T Q^{-1} a$ در تابع هدف می‌باشد.

مثال ۴: مسألهٔ مینیمم کردن مجموع نُرم‌ها [۱]:

$$\min \sum_{i=1}^{m} ||A_i x + b_i||.$$

معادل است با

$$\begin{aligned} \min \quad & \sum_{i=1}^{m} v_{i1} \\ & A_i x + b_i = \bar{v}_i \\ & v_i \in Q_{n_i}, \end{aligned}$$

که $v_i = (v_{i1}; \bar{v}_i)$. به وضوح به شکل یک مسألهٔ برنامه‌ریزی روی مخروط درجه دوم است.

مثال ۵: مسألهٔ مینیمم‌سازی ماکسیمم نرم بردارها [۱]:

$$\min_{x} \max_{1 \le i \le r} ||A_i x + b_i||$$

معادل است با:

$$\begin{aligned} \min \quad & t \\ & A_i x + b_i = v_i \\ & (t; v_i) \in Q_{n_i}, \forall i = 1, ..., r. \end{aligned}$$

مثال ۶: مسألهٔ مینیمم‌سازی میانگین توافقی توابع آفین مثبت [۱]:

$$\min \sum_{i=1}^{r} \frac{1}{a_i^T x + \beta_i}$$

را می‌توان به صورت زیر بیان نمود:

$$\begin{aligned} & \min \sum_{i=1}^{r} u_i \\ & v_i = a_i^t x + \beta_i \\ & u_i v_i \ge 1 \\ & u_i \ge 0. \end{aligned}$$

خود این مسأله معادل است با مسألهٔ برنامه‌ریزی روی مخروط درجه دوم زیر:

$$\min \sum_{i=1}^{r} u_i$$

$$v_i = a_i^T x + \beta_i$$
$$u_i \geq ۰, \ \left(\frac{u_i + v_i}{۲}; \frac{u_i - v_i}{۲}; ۱\right) \in Q_۳.$$

مثال ۷: فرض کنید در یک مسألهٔ برنامه‌ریزی خطی ماتریس ضرایب قیدها، A، با خطای جزئی در دست باشند. به عبارت دیگر، مسأله [۳، ۴]:

$$\begin{aligned} \min \quad & c^T x \\ & a_i^T x \leq b_i, \ \forall i = ۱, ..., m, \\ & a_i \in E_i = \{\bar{a}_i + P_i u | \ \ ||u|| \leq ۱\}, \forall i = ۱, ..., m, \end{aligned}$$

را در نظر بگیرید، که در اینجا E_i نمایانگر یک بیضی، P_i نیز یک ماتریس $n \times n$ و $\bar{a}_i$ نیز مقدار معلوم و نادقیق a_i است. قید خطی $a_i{}^T x \leq b_i, \ a_i \in E_i$ را به صورت $\sup\{a_i^T x | a_i \in E_i\} \leq b_i$ می‌نویسیم که این خود معادل است با:

$$\bar{a}_i{}^T x + ||P_i^T x|| \leq b_i.$$

به وضوح این قید به شکل مخروط درجه دوم است. بنابراین مسألهٔ برنامه‌ریزی خطی با خطای بیضوی در ماتریس ضرایب را می‌توان به شکل یک برنامه‌ریزی روی مخروط درجه دوم بیان کرد.

لازم به ذکراست که مسائل فراوانی به صورت برنامه‌ریزی روی مخروط درجه دوم قابل بیان هستند که خواننده می‌تواند برای کسب اطلاعات بیشتر به [۱، ۳، ۴] مراجعه کند.

۴. خواص جبری مخروط‌های درجه دوم و شرایط بهینگی

برای بیان شرایط بهینگی برنامه‌ریزی روی مخروط‌های درجه دوم از جبر خاصی موسوم به جبر جردن استفاده می‌شود [٦]. در ادامه ابتدا به معرفی این جبر و برخی خواص آن می‌پردازیم، سپس قضایای ضعیف و قوی دوگانی و شرایط بهینگی برنامه‌ریزی روی مخروط را بیان و اثبات می‌کنیم [۱، ۳]. مشابه بخش قبل برای سادگی فرض می‌کنیم که مسأله شامل فقط یک مخروط درجه دوم است.

تعریف ۱.۴. حاصل ضرب xoy دو بردار $x = (x_۱; \bar{x}) \in R^n, \ y = (y_۱; \bar{y}) \in R^n$، که $x_۱, y_۱ \in R$ و $\bar{x}, \bar{y} \in R^{n-۱}$ را به صورت زیر تعریف می‌کنیم

$$y = (y_۱, y_۲, y_۳)^t$$
$$x = (x_۱, x_۲, x_۲)^t$$

که

$$xoy = (x^T y, x_۱ y_۲ + y_۱ x_۲, ..., x_۱ y_n + y_۱ x_n)^T = Arw(x) y,$$

که

$$Arw(x) = \begin{pmatrix} x_1 & \bar{x}^T \\ \bar{x} & x_1 I \end{pmatrix}.$$

مثلاً اگر $x = (x_1, x_2, x_3)^t$ و $y = (y_1, y_2, y_3)^t$ آنگاه

$$xoy = \begin{bmatrix} x_1 & x_2 & x_3 \\ x_2 & x_1 & 0 \\ x_3 & 0 & x_1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

ساختار $(\mathbb{R}^n, o)$ را جبر جردن بر $\mathbb{R}^n$ می‌نامیم.
برخی خواص جبر (R^n, o) به شرح زیر است.

- $xo(\alpha y + \beta z) = \alpha xoy + \beta xoz$ برای هر $\alpha, \beta \in R$ و $x, y, z \in R^n$.
- $xoy = yox$ برای هر $x, y \in R^n$.
- بردار $e = (1, 0, 0, ..., 0)^T$ بردار یکه (R^n, o) است، یعنی $eox = xoe$ برای هر $x \in \mathbb{R}$.
- جبر جردن شرکت پذیر نیست، در حالی که شرکت‌پذیر توانی است، یعنی $xo(xox) = (xox)ox$. برای سادگی ترکیب p مرتبه از بردار x به‌وسیلهٔ عمل o را با x^p نمایش می‌دهیم. داریم $x^p o x^q = x^{p+q}$.

سه خاصیت اول به آسانی قابل اثباتند، بنابراین در اینجا به اثبات خاصیت چهارم می‌پردازیم. برای این که نشان دهیم o شرکت‌پذیر نیست کافیست یک مثال ارائه دهیم. فرض کنید $x = (3, 1, 2)^T$، $y = (3, -1, -2)^T$ و $z = (1, 2, 1)^T$. به آسانی می‌توان دید که $xo(yoz) \neq (xoy)oz$. بنابراین o شرکت‌پذیر نیست. اثبات شرکت‌پذیری توانی این جبر نیازمند تعریف تجزیهٔ طیفی می‌باشد که خواننده می‌تواند برای مشاهدهٔ جزئیات آن به [1] مراجعه کند.

همچنین لازم است ذکر شود که مخروط درجه دوم نسبت به جبر جردن بسته نیست. به عنوان مثال اگر $x = (1.5, 1, 1)^T$ و $y = (1.5, 1, -1)^T$. آنگاه $x, y \in Q_3$، اما $xoy \notin Q_3$.

در ادامه به بیان شرایط بهینگی مسألهٔ برنامه‌ریزی روی مخروط می‌پردازیم و تفاوت‌های آن را با برنامه‌ریزی خطی بیان می‌کنیم. ابتدا قضیهٔ ضعیف دوگانی را بیان می‌کنیم.

قضیهٔ ۱.۴. (قضیهٔ ضعیف دوگانی) اگر x یک جواب شدنی دلخواه برای مسألهٔ اولیه (۱) و (y, z) یک جواب شدنی دلخواه برای مسأله دوگان (۲) باشد، آنگاه $c^T x \geq b^T y$.

اثبات. مشابه برنامه‌ریزی خطی است.

تعریف ۲.۴ گوییم $(x_1; \bar{x}) \in intQ_n$ اگر $||\bar{x}|| < x_1$.

قبل از این که به بیان قضیهٔ قوی دوگانی بپردازیم، به مثال زیر توجه کنید:

مثال ۸: دوگان مسأله

$$\begin{aligned} \min \quad & (1,-1,0)x \\ & (0,0,1)x = 1 \\ & x \in Q_3. \end{aligned}$$

در $\mathbb{R}^3$، به صورت زیر است:

$$\begin{aligned} \max \quad & (0,0,1)y \\ & \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} y + z = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \\ & z \in Q_3. \end{aligned}$$

قید مسأله اولیه نتیجه می‌دهد که $x_3 = 1$، در نتیجه این مسأله به صورت زیر در می‌آید:

$$\begin{aligned} \min \quad & x_1 - x_2 \\ & x_1 \geq \sqrt{x_2^2 + 1}. \end{aligned}$$

از قید این مسأله نتیجه می‌شود که $x_1 - x_2 > 0$ و هر اندازه که بخواهیم می‌تواند به صفر نزدیک شود ولی صفر نمی‌شود. از قید مسأله دوگان داریم $z^T = (1,-1,-y)^T$. بنابراین می‌توان آن را به صورت زیر نوشت:

$$\begin{aligned} \max \quad & y \\ & 1 \geq \sqrt{1 + y^2}, \end{aligned}$$

که این خود نتیجه می‌دهد $y = 0$. بنابراین مسألهٔ دوگان جواب بهینه یکتای خود را که همان صفر است می‌گیرد در حالی که اگرچه مسألهٔ اولیه جواب دارد اما شدنی است. به عبارت دیگر، در مسألهٔ برنامه‌ریزی خطی شکاف دوگانی[1] در صورتی که هر دو مسأله اولیه و دوگان شدنی باشند صفر است در حالی که این شکاف در برنامه‌ریزی روی مخروط درجه دوم لزوماً صفر نیست. یکی از این مسائل می‌تواند جواب بهینه داشته باشد و مسألهٔ دوم می‌تواند از این ویژگی بهره‌مند نباشد. در مثال زیر برعکس مثال قبلی مسألهٔ اولیه دارای جواب بهینه است ولی دوگان آن نشدنی می‌باشد:

مثال ۹: مسألهٔ

$$\begin{aligned} \min \quad & x_2 \\ & x_1 - x_3 = 0 \\ & x \in Q_3. \end{aligned}$$

1) Duality gap

در $\mathbb{R}^3$ معادل است با

$$\begin{aligned} \min \quad & x_2 \\ & x_2 = 0 \\ & x_1 \geq 0 \end{aligned}$$

که مقدار بهینه آن نیز به وضوح صفر است. حال دوگان این مسأله، یعنی

$$\begin{aligned} \max \quad & 0 \\ & y + z_1 = 0 \\ & z_2 = 1 \\ & -y + z_3 = 0 \\ & z \in Q_3 \end{aligned}$$

شدنی نیست. این نیز یک نمونه از ضعف های نظریهٔ دوگانی در برنامه‌ریزی روی مخروط درجه دوم در مقایسه با برنامه‌ریزی خطی می‌باشد. دو مثال فوق نشان می دهند که برای داشتن نتایجی مشابه برنامه‌ریزی خطی نیازمند فرضیات قویتری روی مسائل اولیه و دوگان هستیم.

قضیهٔ زیر شرایطی را بیان می‌کند که تحت آن‌ها هر دو مسألهٔ اولیه و دوگان جواب بهینه دارند و مقدار تابع هدف آنها نیز یکی است.

قضیهٔ ۲.۴ (قضیه قوی دوگانی) اگر هر دو مسألهٔ اولیه و دوگان نقاط شدنی درونی داشته باشند، آنگاه هر دو جواب‌های بهینه دارند و مقدار تابع هدفشان نیز برابر می‌باشد.

اثبات: روند اثبات مشابه حالت برنامه‌ریزی خطی می‌باشد و به دلیل طولانی بودن آن در اینجا از آوردن آن صرف نظر می‌کنیم [۳].

واضح است که قضیهٔ فوق ضعیف‌تر از قضیهٔ قوی دوگانی در حالت برنامه‌ریزی خطی می‌باشد. زیرا در برنامه‌ریزی خطی نیازی به این که مسائل اولیه و دوگان دارای نقاط شدنی درونی باشند نیست. درحالی که همانطور که در مثال‌های ۸ و ۹ دیدیم در برنامه‌ریزی روی مخروط درجه دوم بدون این شرط نمی‌توان برابر بودن توابع هدف مسائل اولیه و دوگان را در حالت کلی نتیجه گرفت.

در قضیهٔ زیر به بیان شرایط بهینگی می‌پردازیم.

قضیهٔ ۳.۴ (شرایط بهینگی) اگر مسائل اولیه و دوگان به ترتیب دارای جواب‌های شدنی درونی x و (y,z) باشند، آنگاه x و (y,z) جواب‌های بهینه هستند اگر و فقط اگر

$$\begin{aligned} & Ax = b,\ x \in Q \\ & A^T y + z = c,\ z \in Q \qquad (3) \\ & x o z = 0. \end{aligned}$$

اثبات: اگر برای نقاط اکیداً شدنی x و (y,z) شرایط (۳) برقرار باشند، آنگاه از $xoz = \circ$ نتیجه می‌شود که $x^Tz = \circ$ یا $c^Tx = b^Ty$ که این خود بیانگر این است که نقاط داده شده جواب‌های بهینه برای مسائل اولیه و دوگان هستند. برای اثبات عکس قضیه فرض کنید که x و (y,z) نقاط شدنی درونی و بهینه برای مسائل اولیه و دوگان باشند. در این صورت دو شرط اول (۳) برقرارند. برای درستی شرط آخر با توجه به $x^Tz = \circ$ داریم: $||\bar{x}||||\bar{z}|| \leq x_\backslash z_\backslash = -\bar{x}^T\bar{z} \leq ||\bar{z}||||\bar{x}||$. بنابراین یا $x,z = \circ$ که در این صورت حکم قضیه واضح است و یا $\bar{x} = -\alpha\bar{z}$ برای $\alpha > \circ$، $||\bar{x}|| = x_\backslash$ و $||\bar{z}|| = z_\backslash$. در این حالت نیز به وضوح $xoz = \circ$.

توجه ۱.۴ در این مقاله برای سادگی در بیان قضایا و مفاهیم فرض کردیم که در مسألهٔ برنامه‌ریزی روی مخروط درجه دوم فقط یک مخروط داریم. در حالی که مسأله می‌تواند دارای چندین مخروط درجه دوم از ابعاد مختلف باشد. مثلاً مسألهٔ:

$$\begin{aligned}\min \quad & c_\backslash x_\backslash + \ldots + c_k x_k \\ & A_\backslash x_\backslash + \ldots + A_k x_k = b \\ & x_i \in Q_{n_i}, \quad \forall i = \backslash, \ldots, k,\end{aligned}$$

و دوگان آن

$$\begin{aligned}\max \quad & b^Ty \\ & A_i^T y_i + z_i = c_i, \forall i = \backslash, \ldots, k, \\ & z_i \in Q_{n_i},\end{aligned}$$

که Q_{n_i} ها مخروط‌های درجه دوم با ابعاد متفاوت‌اند، از این گونه مسائل‌اند. در نهایت شرایط بهینگی برای این مسأله نیز به صورت زیر می‌باشد:

$$\begin{aligned}& A_\backslash x_\backslash + \ldots + A_k x_k = b, \; x_i \in Q_{n_i}, \forall i = \backslash, \ldots, k, \\ & A_i y_i + z_i = c_i, \; z_i \in Q_{n_i}, \quad \forall i = \backslash, \ldots, k, \\ & x_i o z_i = \circ, \forall i = \backslash, \ldots, k.\end{aligned}$$

توجه ۲.۴ برای حل مسائل برنامه‌ریزی روی مخروط‌های درجه دوم نرم‌افزارهای آموزشی و تجاری متعددی طراحی شده است. به عنوان مثال، $LOQO$، $SeDumi$ و $Mosek$ چند نمونه از آنها هستند. نرم‌افزار $LOQO$ از روش‌های مانعی لگاریتمی[1] و تکنیک‌های نوین در هموار نمودن قیدهای ناهموار [۹] استفاده می‌کند در حالی که $SeDumi$ و $Mosek$ از روش‌های نقطه درونی و شرایط بهینگی به‌دست آمده و از جبر جردن [۲] استفاده می‌کنند.

1) Logarithmic Barrier Methods

مراجع

[1] F. Alizadeh and D. Goldfarb, Second order cone programming, *Mathematical Programming*, 95(1), 3–51, 2003.

[2] E. D. Andersen and K.D. Andersen. The MOSEK interior point optimizer for linear programming: an implementation of the homogeneous algorithm. In: H. Frenk, K. Roos, T. Terlaky, and S. Zhang, editors, *High Performance Optimization*, Kluwer Academic Publishers, 197–232, 2000.

[3] A. Ben-Tal and A. Nemirovski, Lectures on Modern Convex Optimization. Analysis, Algorithms, and Engineering Applications, SIAM, 2001.

[4] S. Boyd and L. Vandenberghe,*Convex Optimization,* Cambridge University Press, 2004.

[5] G. B. Dantzig, *Linear Programming and Extensions,* Princeton University Press, 1963.

[6] J. Faraut and A. Korányi, *Analysis on Symmetric Cones,* Oxford University Press, Oxford, UK, 1994.

[7] R. Fletcher, *Practical Methods of Optimization,* John Wiley, Chichester, 1980.

[8] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, Applications of second order cone programming, *Linear Algebra Appl.*,284, 193– 228, 1998.

[9] R. J. Vanderbei and D. Shanno, Using LOQO to solve second order cone programming problems, Technical Report SOR-98-09, Statistics and Operations Research, Princeton University, 1998.

[10] G. O. Wesolowsky, The Weber problem: history and perspectives, Location Sci. 1, 5–23, 1993.

مازیار صلاحی

گروه ریاضی، دانشگاه گیلان

salahim@guilan.ac.ir

فرهنگ و اندیشهٔ ریاضی
شماره ۴۱ (پاییز ۱۳۸۷) صص. ۳۹ تا ۴۲

دربارهٔ زیرگروه‌هایی با شاخص اول

تی. وای. لَم*

ترجمهٔ حمیدرضا وهابی

یک بخش ثابت در هر درس مقدماتی جبر مجرد یا نظریهٔ گروه‌ها این حقیقت است که هر گاه G یک گروه باشد، هر زیرگروه H با شاخص ۲ در G نرمال است. به عنوان تعمیمی از این مطلب (و نشان دادن این که واقعاً عدد اول ۲ در این جا عدد اول ویژه‌ای نیست)، برخی از مدرسین قضیهٔ جالب زیر را بیان می‌کنند.

قضیهٔ ۱. اگر H یک زیرگروه از گروه متناهی G باشد به طوری که شاخص $[G:H]$ کوچکترین عدد اوّلی باشد که $|G|$ را عاد می‌کند، آن‌گاه H در G نرمال است.

برهان این قضیه معمولاً بر این حقیقت استوار است که عمل ضرب متعدی G روی فضای همدسته‌های چپ H، هسته‌ای مانند K دارد که بزرگترین زیرگروه نرمال G مشمول در H است. این مطلب نشان می‌دهد که $[G:K] = [G:H][H:K]$ بر $[G:H]!$ بخش پذیر است. سپس فرض روی $[G:H]$ نتیجه می‌دهد که $[H:K] = ۱$، و از این رو $H = K$ در G نرمال است. (برای مثال نگاه کنید به [۱، ص ۱۲۳]، [۲، ص ۴۴].) این برهان هرچند کوتاه، اما برای دانشجویان مبتدی مناسب نیست زیرا برنامه‌های درسی آنها شامل نظریهٔ عمل گروه‌ها نیست. هنگام تدریس یک درس جبر مجرد مقدماتی، برهانی برای قضیهٔ ۱ یافتم که هم سریع است و هم درک آن برای دانشجویانی که چنین درسی را می‌گذرانند آسان. به علاوه، ایدهٔ این برهان باعث شد که محک‌هایی برای نرمال بودن زیرگروه‌هایی با شاخص اوّلِ گروه دلخواه G پیدا کنم. فکر نمی‌کنم که چنین محک‌هایی در متون استاندارد یا حتی کتاب‌های عمومی نظریهٔ گروه‌ها یافت شود، از این رو آنها را در این جا آورده‌ام.استدلالی را که می‌آورم به نوعی مشابه برهان معمولی

1) T. Y. Lam

این مطلب است که زیرگروهِ یک گروه دوری، دوری است؛ و به طور قطع دانشجویان چنین برهانی را در برنامه‌های درسی خود دیده‌اند.

قضیهٔ ۲. هر گاه G یک گروه و H یک زیرگروه G با شاخص اول p باشد، آن‌گاه گزاره‌های زیر معادل‌اند:

۱− H در G نرمال است،

۲− برای هر $a \in G - H$، $a^p \in H$،

۳− برای هر $a \in G - H$، $a^n \in H$ به ازای عدد صحیح مثبتی مانند $n = n(a)$ که هیچ عامل اولی کوچکتر از p ندارد.

۴− برای هر $a \in G - H$، $a^{۲}, \ldots, a^{p-۱} \notin H$.

اگر G و H در فرض قضیهٔ ۱ صدق کنند، آن‌گاه با توجه به قضیهٔ لاگرانژ، شرط (۳) از قضیهٔ ۲ با انتخاب $n(a) = |G|$ برای هر a در $G - H$ برقرار می‌شود. به هر حال، در قضیهٔ ۲، در حالتی که $|G|$ متناهی است لزومی ندارد که p کوچک‌ترین عامل اول $|G|$ باشد. در واقع p می‌تواند هر عدد اولی باشد که شاخص زیرگروهی مانند H از گروه G است. برای مثال، اگر G گروه متناوب $A_۴$ و H زیرگروه $\{e, (۱۲)(۳۴), (۱۳)(۲۴), (۱۴)(۲۳)\}$ با شاخص $p = ۳$ باشد، در این صورت شرط (۳) از قضیهٔ ۲ با انتخاب $n(a) = ۳$ برای هر a در $G - H$ برقرار می‌شود (زیرا هر یک از این a ها دوری به طول ۳ هستند). از این رو می‌توانیم نرمال بودن H را از قضیهٔ ۲ نتیجه بگیریم. از طرف دیگر، اگر G گروه متقارن $S_۴$ و H نسخه‌ای از گروه دو وجهی هشت عضوی (گروه تقارن‌های مربع به عنوان زیرگروهی از $S_۴$) با $p = [G : H] = ۳$ باشد، به وضوح شرایط (۲) و (۴) از قضیهٔ ۲ برقرار نیستند (با در نظر گرفتن یک ترانهش a در $G - H$)، پس نتیجه می‌گیریم که H در G نرمال نیست.

برهان قضیهٔ ۲. نشان می‌دهیم که $(۱) \Rightarrow (۴) \Rightarrow (۳) \Rightarrow (۲) \Rightarrow (۱)$.

$(۲) \Rightarrow (۱)$. بنابر فرض (۱)، گروه خارج قسمتی $\frac{G}{H}$ را می‌توان ساخت که در آن $|\frac{G}{H}| = p$، در نتیجه شرط (۲) به دست می‌آید (مثلاً با استفاده از قضیهٔ لاگرانژ).

$(۳) \Rightarrow (۲)$. این قسمت با انتخاب $n(a) = p$ برای هر a در $G - H$ واضح است (توجه کنید که p یک عدد اوّل است).

$(۴) \Rightarrow (۳)$. عضو دلخواه a را در $G - H$ در نظر بگیرید و فرض کنید $n = n(a)$ همان باشد که در شرط (۳) بیان شد. فرض کنید s کوچکترین عدد صحیحی باشد که $a^s \in H$. می‌توانیم بنویسیم $n = qs + r$ که در آن q و r اعداد صحیح‌اند و $۰ \leq r < s$. بنا به فرض، H شامل $a^n = (a^s)^q.a^r$ است، بنابراین $a^r \in H$. این نتیجه می‌دهد که $r = ۰$، پس $n = qs$. چون به وضوح $s > ۱$، فرض روی $n = n(a)$ در (۳) نتیجه می‌دهد $s \geq p$، که شرط (۴) را ثابت می‌کند.

(۱) $\Rightarrow$ (۴). فرض کنید شرط (۴) برقرار بوده ولی H در G نرمال نباشد. در این صورت $b \in G$ و $h \in H$ وجود دارند به طوری که $a := bhb^{-1} \notin H$. از آن جا که $xH = yH$ اگر و فقط اگر $x^{-1}y \in H$، (۴) نتیجه می‌دهد که

$(*)$ $\quad H, aH, \cdots, a^{p-1}H$، p همدستهٔ چپ مجزای H در G هستند.

از $bh = ab$، نتیجه می‌گیریم $bH = abH$. واضح است که $b \notin H$. پس بنابر $(*)$، $bH = a^iH$ به ازای iیی که $1 \leq i \leq p-1$. اما در این صورت $a^iH = bH = abH$، و حذف a نتیجه می‌دهد که $a^{i-1}H = bH = a^iH$، که با $(*)$ در تناقض است.

تبصره‌ها.

الف. شایان توجه است که در برهان قضیهٔ ۲، از فرض اوّل بودن p فقط در استنتاج بدیهی (۳) $\Rightarrow$ (۲) استفاده شده است. بنابراین، برهان‌هایی که در سه استنتاج دیگر استفاده کردیم برای هر شاخص (متناهی) $p = [G : H]$ معتبر است. این بدان معناست که برای هر شاخص متناهی p، شرط (۲) هم‌چنان یک شرط لازم و هر کدام از شرایط (۳) و (۴) هم‌چنان یک شرط کافی برای نرمال بودن H در G است. ولی در حالتی که p اوّل نباشد نه (۳) و نه (۴) شرط لازم برای نرمال بودن H نیست. برای مشاهدهٔ این مطلب، G را یک گروه دوری از مرتبهٔ مربع یک عدد اول و H را $\{e\}$ بگیرید. به طور مشابه در حالتی که p اوّل نباشد، (۲) یک شرط کافی برای نرمال بودن H نخواهد بود. برای مشاهدهٔ این مطلب G را یکی از دو گروه غیر آبلی از مرتبهٔ l^3 (برای عدد اول فرد l) و H را زیرگروه تولیدشده به‌وسیلهٔ یک عضو غیرمرکزی از مرتبهٔ l بگیرید.

ب. در حالتی که شاخص یک عدد اوّل است، تام دورسی[1] به مؤلف متذکر شده است که شرایط (۲)، (۳) و (۴) در قضیهٔ ۲ همچنان معتبر خواهند بود اگر به جای واژه‌های «برای هر» در آنها واژه‌های «برای برخی» بنویسیم. برهان این نکتهٔ جالب را به عنوان تمرینی به عهدهٔ دانشجویان می‌گذاریم.

برای مشاهدهٔ امکان کاربرد قضیهٔ ۲ در مورد گروه‌های نامتناهی، نتیجهٔ مشهور زیر را ثابت می‌کنیم.

نتیجه. اگر p یک عدد اوّل و مرتبهٔ هر عضو گروه G توانی از p باشد، آن گاه هر زیرگروه H با شاخص p در G نرمال است.

برهان. برای هر a در $G - H$، می‌توانیم $n(a)$ را مرتبهٔ a بگیریم، پس شرط (۳) در قضیهٔ ۲ برقرار است.

1) Tom Dorsey

مراجع

[1] D. S. Dummit and R. M. Foote, *Abstract Algebra*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, 1999.

[2] I. M. Isaacs, Algebra, Brooks-Cole, Pacific Grove, CA, 1994.

مترجم: حمیدرضا وهابی
دانشگاه آزاد اسلامی واحد اسلامشهر
hrvahabi@yahoo.com

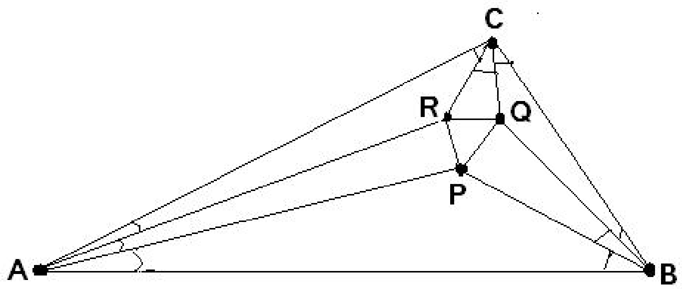
فرهنگ و اندیشهٔ ریاضی
شماره ۴۱ (پاییز ۱۳۸۷) صص. ۴۳ تا ۵۲

اثبات جدیدی از قضیه مورلی

محمدرضا درفشه

قضیهٔ مورلی

نقاط برخورد خطوط مجاور اضلاع تثلیث کنندهٔ سه زاویهٔ داخلی هر مثلث تشکیل یک مثلث متساوی‌الاضلاع می‌دهد.



(شکل ۱)

مسألهٔ بالا در ۱۸۹۹ توسط فرانک مورلی که در آن موقع استاد ریاضی یکی از دانشگاه‌های آمریکا بود مطرح گردید. در مقاله ۱۹۲۹ [7] مورلی برای اولین بار از این قضیه نام می‌برد و می‌نویسد که در مورد ویژگی‌های کاردیوئیدها تحقیق می‌کرده است که به این مسأله برخورد می‌نماید.*

این مسأله در بخش مسائل ریاضی و حل آنها در [5] در ۱۹۰۹ مطرح شد. یکی از روش‌ها که توسط شخصی به نام ام. ت. نارانین‌گار[1] ارسال شده بود، مورد توجه کاکستر و گرتیزر قرار گرفت [4].

*) خوانندهٔ علاقمند برای دیدن تعریف کاردیوئید می‌تواند به صفحهٔ ۲۳۱ جلد اول [۱] مراجعه نماید، بخصوص رابطهٔ موجود مثلث مورلی و کاردیوئیدها که در صفحهٔ ۲۷۶ همین مرجع آورده شده است.

1) M. T. Naraniengar

به این ترتیب این روش حل معروف شد. از آن سال به بعد این مسألهٔ جالب وارد فولکور ریاضی شد و طبق معمول علاقمندان به ریاضیات، به‌خصوص آن‌ها که دوست‌دار هندسه هستند، کوشش می‌کنند که بتوانند یک راه ساده و مقدماتی و قابل فهم برای کسانی که اطلاعاتی از هندسهٔ مسطحه و به‌خصوص ویژگی‌های مثلث‌ها دارند بیابند. بیشتر اثبات‌هائی که تا سال ۱۹۹۸ میلادی برای قضیهٔ مورلی ارائه گردید هندسی یا مثلثاتی بودند [4] و [1] و [8]. در [6] راه حلی ارائه شده که از نمایش اعداد مختلط در صفحه استفاده شده است. این مسأله در ۱۹۹۸ به الن. کُن، برندهٔ جایزهٔ فیلدز در ریاضیات در ۱۹۸۲ میلادی، معرفی گردید و وی توانست یک راه حل فوق‌العاده زیبا که دارای طبیعت هندسی ــ جبری است ارائه کند. راه حل کُن به‌صورت مقاله‌ای به مناسبت چهلمین سالگرد تأسیس مؤسسهٔ تحقیقاتی IHES در مجموعه مقالات چاپ گردید [2].

لازم به توضیح است که استفاده از اعداد مختلط در صفحه برای حل مسائل هندسهٔ مسطحه موضوع جدیدی نیست ولی استفادهٔ هوشمندانهٔ کُن از این روش در حل مسألهٔ مثلث مورلی نوآوری محسوب می‌شود. آلن کُن خود نیز مجدداً در ۲۰۰۴ مقاله‌ای تحت عنوان تقارن، در خبرنامهٔ انجمن ریاضی اروپا منتشر نمود و در پایان مقاله به خطوط کلی اثبات جدید خود از قضیهٔ مورلی پرداخت [3]. همین مقاله به بخش مسابقات کمیتهٔ عمومی‌سازی ریاضیات انجمن ریاضی اروپا تسلیم گردید و رتبهٔ دوم را کسب نمود.

آنچه که مایهٔ کنجکاوی افرادی چون من قرار می‌گیرد این است که چگونه چنین مسأله‌ای مقدماتی توجه ریاضیدانی چون کن که در ریاضیات تحقیقات ارزنده و نیز پژوهش‌های عمیقی در هندسهٔ ناجابجائی دارد، به خود جلب می‌کند و آنچه بیشتر مایهٔ تعجب من می‌شود این است که چگونه است که قضیه‌ای که معروفیتش تقریباً به اندازهٔ قضیهٔ فیثاغورث است از چشم ریاضیدانی نظیر کن به دور مانده بخصوص ظاهراً در زمان نوجوانی وی هنوز شعار «مرگ بر هندسهٔ اقلیدسی» در فرانسه داده نشده بود و هندسهٔ اقلیدسی در برنامهٔ درسی مدارس فرانسه از جایگاه والایی برخوردار بوده است. نکتهٔ جالب‌تر این که این روش حل ناشی از طرز تفکر هندسی و جبری کن است که در واقع رشتهٔ تخصصی ایشان می‌باشد. پس از خواندن راه حل کن با خودم زمزمه کردم خوشا به حال اشخاصی که هم با ایده‌های بزرگ ریاضی و هم با مسائل کوچک ریاضی خستگی خود را در می‌کنند.

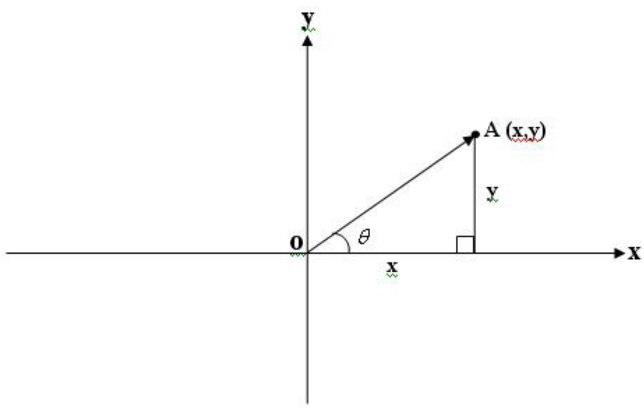
راه حل کن مبتنی بر نمایش اعداد مختلط در صفحه به عنوان نقاط و تبدیلات آفین صفحهٔ مختلط است. در این نوشته این روش حل را تا حدودی بسط داده‌ام به طوری که برای تعداد زیادتری از خوانندگان قابل فهم باشد، معهذا اطلاعاتی مقدماتی دربارهٔ گروه و میدان مورد نیاز هستند. اما قبل از پرداختن به مقدمات اثبات توضیح این مطلب ضروری است که قضیهٔ مورلی در مجلات ریاضی که به زبان فارسی منتشر می‌شوند نیز مورد توجه قرار گرفته است. شادروان استاد حسین غیور در مقالاتی که در مجلهٔ رشد ریاضی به چاپ رساند [9] و [10] به روش‌های اثبات قضیهٔ مورلی و تعمیم این قضیه پرداخت. بخصوص که وی اثباتی ارائه می‌دهد که مربوط به ایجاد

مثلث‌های مورلی حاصل از سه قسمت کردن زوایای خارجی مثلث دلخواه است. مقالهٔ [10] با حوصله و دقت فراوان نگاشته شده و خواندن آن به علاقمندان توصیه می‌گردد.

صفحهٔ مختلط همان $\mathbb{C} = \mathbb{R} \times \mathbb{R} = \mathbb{R}^۲$ است، که هر عضو (x, y) آن را می‌توان به عنوان نقطهٔ A با مختصات (x, y) در صفحه نمایش داد. اگر $(۰ \text{ و} ۱) = ۱$ نقطه‌ای روی محور xها و $i = (۰, ۱)$ نقطه‌ای روی محور yها فرض شود، آنگاه می‌توان نوشت

$$(x, y) = (x, ۰) + (y, ۰) = x(۱, ۰) + y(۰, ۱) = x + iy$$

در این‌صورت می‌نویسیم $Z_A = x + iy$. واضح است که $A \to Z_A$ یک تناظر یک به یک بین نقاط صفحه و اعداد مختلط است. از این رو، هر دو مفهوم را یکی می‌گیریم. اگر $Z_A = x + iy$ فرض شود آنگاه بردار $\overrightarrow{OA}$ را به عنوان برداری به مبداء O (مبداء مختصات) و انتهای A مطابق شکل ۲ در نظر می‌گیریم:



(شکل ۲)

طول بردار $\overrightarrow{OA}$ برابراست با $\rho = \sqrt{x^۲ + y^۲}$ ، $\sin\theta = \dfrac{y}{\rho}$ ، $\cos\theta = \dfrac{x}{\rho}$ و θ زاویه‌ایست که $\overrightarrow{OA}$ با جهت مثبت محور xها می‌سازد. در نتیجه می‌توان نوشت

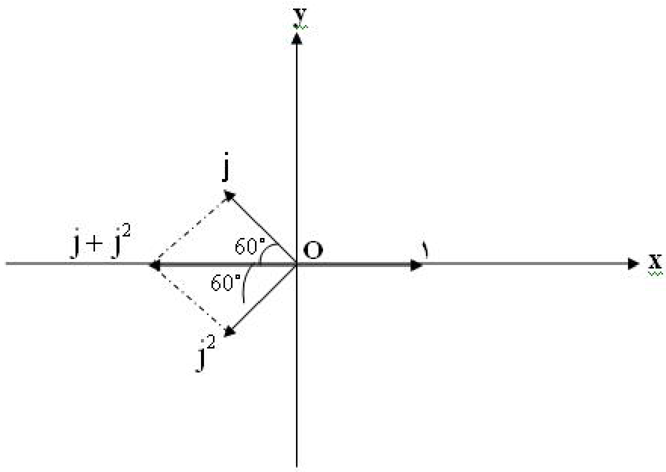
$$Z_A = x + iy = \rho(\frac{x}{\rho} + i\frac{y}{\rho}) = \rho(\cos\theta + i\sin\theta) = \rho e^{i\theta}$$

$Z_A = \rho e^{i\theta}$ که صورت قطبی نقطهٔ A نامیده می‌شود و به طور منحصر به فردی توسط $\rho = \sqrt{x^۲ + y^۲}$ و $\tan\theta = \frac{y}{x}$ به دست می آید.

جمع و تفریق بردارها مطابق معمول انجام می‌گیرد. یعنی اگر $Z_A = x_۱ + iy_۱$ و $Z_B = x_۲ + iy_۲$ مختصات بردارهای $\overrightarrow{OA}$ و $\overrightarrow{OB}$ باشند آنگاه $Z_A + Z_B = x_۱ + x_۲ + i(y_۱ + y_۲)$ و $Z_A - Z_B = x_۱ - x_۲ + i(y_۱ - y_۲)$ و از لحاظ هندسی $Z_A + Z_B$ متناظر است با برداری که قطر متوازی‌الاضلاع با ضلع‌های $\overrightarrow{OA}$ و $\overrightarrow{OB}$ است و $Z_A - Z_B$ متناظر است با بردار $\overrightarrow{BA}$.این

مجموع و تفاضل را به ترتیب با $A+B$ و $A-B$ نمایش می‌دهیم. ضرب بردارها حالت ویژه‌ای دارد. اگر $Z_A = x_1 + iy_1$ و $Z_B = x_2 + iy_2$ آنگاه $Z_A Z_B = x_1x_2 - y_1y_2 + i(x_1y_2 + y_1x_2)$ و لذا مختصات بردار متناظر با $Z_A Z_B$ که با AB نمایش می‌دهیم را می‌توان محاسبه نمود. اگر $Z_A = \rho_1 e^{i\theta_1}$ و $Z_B = \rho_2 e^{i\theta_2}$ مختص‌های قطبی A و B باشند آنگاه $Z_A Z_B = \rho_1\rho_2 e^{i(\theta_1+\theta_2)}$ یعنی $Z_A Z_B$ متناظر است با برداری چون $\overrightarrow{OC}$ که طول آن $\rho_1\rho_2$ و زاویه‌اش با جهت مثبت محور $\overrightarrow{OX}$، $\theta_1 + \theta_2$ است.

اکنون اگر نقطهٔ $j = e^{\frac{2\pi i}{3}}$ در $\mathbb{C} = \mathbb{R}^2$ را در نظر بگیریم آنگاه داریم $j^2 + j + 1 = 0$ که ۱ همان نقطهٔ $(1, 0)$ است. از لحاظ هندسی j برداری است واحد که با جهت مثبت محور $\overrightarrow{OX}$ زاویه $\frac{2\pi}{3}$ می‌سازد و j^2 برداری واحد است که با $\overrightarrow{OX}$ زاویه $\frac{4\pi}{3}$ می‌سازد.



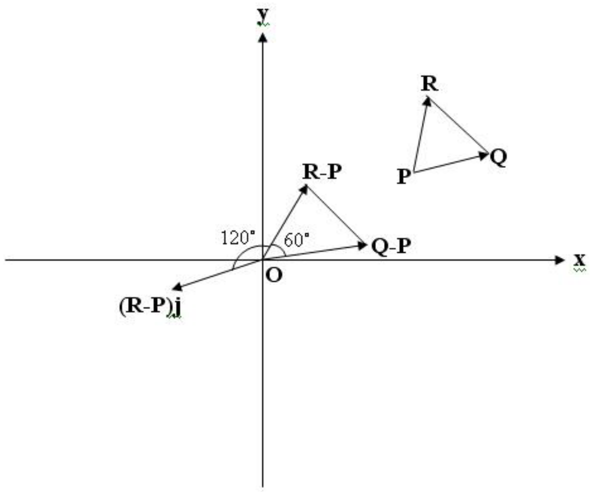
(شکل ۳)

چون j و j^2 با جهت منفی محور xها زاویه $\frac{\pi}{3}$ می‌سازند و طول مساوی دارند پس قطر متوازی‌الاضلاع با اضلاع j و j^2 روی محور xها واقع و قرینهٔ بردار ۱ با طول ۱ است، بنابراین $j^2 + j + 1 = 0$.

اکنون به اثبات یک سرشت‌نمایی کلاسیک مثلث متساوی‌الاضلاع می‌پردازیم که در اینجا مورد استفاده واقع می‌گردد.

لم ۱. مثلث PQR در صفحهٔ $\mathbb{C} = \mathbb{R}^2$ متساوی‌الاضلاع است اگر و تنها اگر $P + Qj + Rj^2 = 0$.

برهان. مطابق شکل ۴ فرض می‌کنیم مثلث PQR متساوی‌الاضلاع است.



(شکل ۴)

$\overrightarrow{OR}-\overrightarrow{OP}$ متناظر است با بردار $\overrightarrow{PR}$ و $\overrightarrow{OQ}-\overrightarrow{OP}$ متناظر است با بردار $\overrightarrow{PQ}$. بنا به توضیحی که در ابتدا دادیم قرار می‌دهیم $\overrightarrow{OR}-\overrightarrow{OP}=R-P$ و $\overrightarrow{OQ}-\overrightarrow{OP}=QP$. اکنون واضح است که مثلث PQR متساوی الاضلاع است اگر و تنها اگر مثلث با رئوس O، $R-P$ و $Q-P$ متساوی‌الاضلاع باشد. بردار $(R-P)j$ را در نظر می‌گیریم. طول این بردار با طول بردار $\overrightarrow{O(R-P)}$ مساوی است و زاویه‌ای که با $\overrightarrow{O(R-P)}$ می‌سازد مساوی است با ۱۲۰° . در نتیجه $\overrightarrow{O(R-P)j}$ و $\overrightarrow{O(Q-P)}$ مساوی‌اند. بنابراین

$$(R-P)j+Q-P=\circ.$$

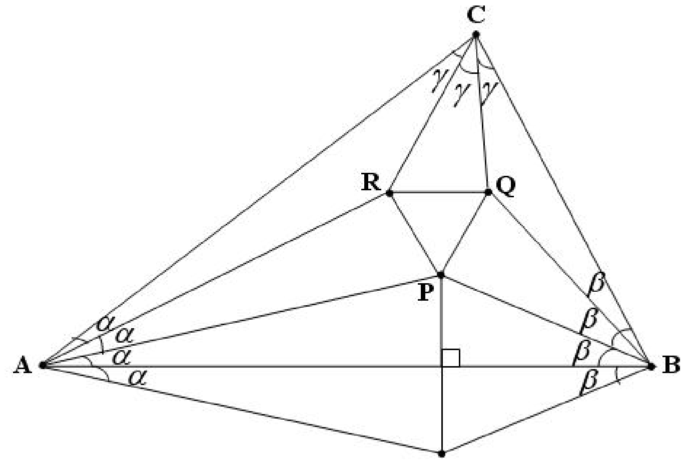
در نتیجه $Rj+Q-P(j+۱)=\circ$ چون $j^۲+j+۱=\circ$، حاصل می‌گردد

$$Rj+Q+Pj^۲=\circ$$

اگر دو طرف تساوی اخیر را از سمت راست در j ضرب کنیم و از تساوی $j^۳=۱$ استفاده نمائیم، حاصل می گردد $P+Qj+Rj^۲=\circ$. بعکس فرض کنید P، Q و R رئوس یک مثلث در صفحه‌اند به طوری که $P+Qj+Rj^۲=\circ$. با استفاده از رابطهٔ $۱+j+j^۲=\circ$ حاصل می‌شود $R-P=(Q-R)j$، چون طول j ۱ است نتیجه می‌شود که طول‌های $R-P=\overrightarrow{PR}$ و $Q-R=\overrightarrow{RQ}$ مساوی‌اند. همچنین با استفاده از $P+Qj+Rj^۲=\circ$ و $j=-۱-j^۲$ حاصل می‌شود $P-Q=(Q-R)j^۲$، چون طول $j^۲$ نیز یک است نتیجه می‌شود که طول‌های $P-Q=\overrightarrow{QP}$ و $Q-R=\overrightarrow{RQ}$ مساوی‌اند. بنابراین طول‌های $\overrightarrow{PR}$، $\overrightarrow{RQ}$ و $\overrightarrow{QP}$ مساوی‌اند و مثلث

PQR متساوی‌الاضلاع و به این ترتیب لم ثابت می‌شود.■

اما اثباتِ کن از قضیهٔ مورلی با این مشاهده آغاز می‌شود که نقاط P، Q و R در شکل ۱ نقاط ثابت حاصلضرب‌های دو به دوی دوران‌های g_1، g_2 و g_3 حول رئوس مثلث‌اند که زاویهٔ دوران در هر مورد مساوی $\frac{2}{3}$ زاویهٔ مثلث می‌باشند. به طور دقیق‌تر اگر g_1، g_2 و g_3 به ترتیب دوران حول رئوس A، B و C با زاویه‌های $\frac{2}{3}A$، $\frac{2}{3}B$ و $\frac{2}{3}C$ باشند آنگاه مثلاً g_1g_2 دورانی است که P را ثابت نگاه می‌دارد. (توجه کنید که ترکیب دوران‌ها نیز مانند ترکیب توابع از سمت چپ به راست نوشته می‌شود یعنی ابتدا g_2 عمل می کند و سپس g_1)



(شکل ۵)

زیرا مطابق شکل ۵ ابتدا g_2 نقطهٔ P را به P' تبدیل می‌کند که چون به اندازهٔ 2β دوران یافته زاویه‌اش با AB باید β باشد و لذا مثلث PBP' و در نتیجه PAP' باید متساوی‌الساقین باشند. بنابراین وقتی g_1 عمل می‌کند P' را به P تبدیل می‌کند در نتیجه P تحت g_1g_2 پایاست. به همین روش ثابت می‌شود Q تحت g_2g_3 و R تحت g_3g_1 پایاست.

برای فرمولبندی مشاهدهٔ کن ناچاریم تبدیلات آفینِ خط حقیقی را تعریف کنیم. از آنجایی که میدان حقیقی $\mathbb{R}$ نقشی در تعریف تبدیل آفین ندارد، لذا آنرا در حالت کلی تعریف می‌کنیم. تبدیلات آفین به طور کلی برای فضای n بعدی قابل تعریف‌اند و آنچه در زیر تعریف می‌شود تبدیل آفین فضای یک بعدی است که به طور خلاصه تبدیل آفین نامیده می‌شود.

تعریف ۱. فرض کنید F یک میدان است. یک تبدیل آفین F عبارتست از نگاشت $f_{a,b}$ ($a, b \in F$)،

($a \neq 0$) با ضابطهٔ زیر

$$f_{a,b} : F \longrightarrow F$$

$$f_{a,b}(x) = ax + b \quad (x \in F).$$

اگر $f_{a,b}$ و $f_{c,d}$ تبدیلات آفین فرض شوند آنگاه به راحتی می‌توان ثابت کرد که $f_{a,b}f_{c,d} = f_{ac,ad+b}$. بنابراین دیده می‌شود که ترکیب دو تبدیل آفین تبدیلی آفین است و هر تبدیل آفین نیز وارونپذیر می‌باشد. این نشان می‌دهد که مجموعهٔ تمام تبدیلات آفین F یک گروه است که آن را با $Af(F)$ نشان می‌دهیم. عضو خنثای این گروه عبارت است از $f_{1,0}$ که با I نمایش می‌دهیم. $c \in F$ را یک نقطهٔ ثابتِ $f_{a,b}$ می‌نامیم هر گاه داشته باشیم $f_{a,b}(c) = c$. در این‌حالت داریم $ac + b = c$ و در نتیجه $c(a-1) = -b$. بنابراین اگر $a \neq 1$ آنگاه $f_{a,b}$ دارای نقطهٔ ثابت منحصر بفرد $c = \frac{b}{1-a}$ است که آن‌را در حالت کلی با $Fix(f_{a,b})$ نمایش می‌دهیم.

نگاشت $d : Af(F) \longrightarrow F^*$ با ضابطه $d(f_{a,b}) = a$ را در نظر می‌گیریم در اینجا F^* گروه ضربی میدان F است. به سادگی می‌توان ثابت کرد که d یک همریختی گروه‌های ضربی است و هستهٔ d عبارتست از $K = \{f_{1,b} | b \in F\}$ که با گروه جمعی F یکریخت است. هر عضو $f_{1,b}$ از K دارای ضابطهٔ $f_{1,b}(x) = x + b$ است که آنرا یک انتقال می‌نامیم و به وضوح دارای نقطهٔ ثابت نمی‌باشد. با استفاده از قضیهٔ اساسی همریختی گروه‌ها داریم

$$Af(F)\diagup_K \cong F^*$$

قضیهٔ زیر در اثبات کن از قضیهٔ مورلی نقش اساسی دارد. نمادگذاری‌های فوق را بدون تکرار در قضیهٔ زیر و اثبات آن به کار می‌بریم.

قضیهٔ ۱. فرض کنید $g_1, g_2, g_3 \in Af(F)$ و g_1g_2، g_2g_3، g_3g_1 و $g_1g_2g_3$ هیچکدام در K نیستند. قرار می‌دهیم $j = d(g_1g_2g_3)$.

در اینصورت شرایط زیر هم ارزند:

(الف) $g_1{}^3 g_2{}^3 g_3{}^3 = I$

(ب) $j^3 = 1$، $j \neq 1$، و $P + Qj + Rj^2 = 0$، که در آن $P = Fix(g_1g_2)$، $Q = Fix(g_2g_3)$ و $R = Fix(g_3g_1)$.

برهان. فرض کنید $g_i = f_{a_i,b_i}$، $1 \leq i \leq 3$. می‌دانیم که $a_i, b_i \in F$ و $a_i \neq 0$. با استفاده از قاعدهٔ ترکیب عناصر $Af(F)$ داریم:

$$g_1g_2 = f_{a_1,b_1}\ f_{a_2,b_2} = f_{a_1a_2,a_1b_2+b_1}$$

$$g_2g_3 = f_{a_2,b_2}\ f_{a_3,b_3} = f_{a_2a_3,a_2b_3+b_2}$$

$$g_3g_1 = f_{a_3,b_3}\ f_{a_1,b_1} = f_{a_3a_1,a_3b_1+b_3}$$

با استفاده از فرمولی که برای نقطهٔ ثابت هر تبدیل آفین یافتیم داریم:

$$P = Fix(g_۱g_۲) = \frac{a_۱a_۲+b_۱}{۱-a_۱a_۲}$$
$$Q = Fix(g_۲g_۳) = \frac{a_۲b_۳+b_۲}{۱-a_۲a_۳}$$
$$R = Fix(g_۳g_۱) = \frac{a_۳b_۱+b_۳}{۱-a_۳a_۱}$$

عبارات فوق بدون ابهام تعریف می‌شوند زیرا فرض کرده‌ایم که هیچکدام از تبدیل‌های $g_۱g_۲$، $g_۲g_۳$ و $g_۳g_۱$ در K نیستند و لذا $a_۱a_۲$، $a_۲a_۳$ و $a_۳a_۱$ همگی مخالف ۱ می‌باشند.
به‌طور مشابه ضابطه‌های $g_۱{}^۳$، $g_۲{}^۳$ و $g_۳{}^۳$ با استفاده از قانون ترکیب در $Af(F)$ محاسبه می‌گردند که عبارتند از:

$$g_۱{}^۳ = f_{a_۱{}^۳,(a_۱{}^۲+a_۱+۱)b_۱},$$
$$g_۲{}^۳ = f_{a_۲{}^۳,(a_۲{}^۲+a_۲+۱)b_۲},$$
$$g_۳{}^۳ = f_{a_۳{}^۳,(a_۳{}^۲+a_۳+۱)b_۳}.$$

و اگر فرض کنیم $g_۱{}^۳g_۲{}^۳g_۳{}^۳ = f_{a,b}$ آنگاه خواهیم داشت:

$$a = a_۱{}^۳a_۲{}^۳a_۳{}^۳$$
$$b = a_۱{}^۳a_۲{}^۳b_۳(a_۳{}^۲ + a_۳ + ۱) + a_۱{}^۳b_۲(a_۲{}^۲ + a_۲ + ۱) + b_۱(a_۱{}^۲ + a_۱ + ۱)$$

اکنون شرایط قضیه را در نظر می‌گیریم. شرط الف برقرار است اگر و تنها اگر $a = ۱$ و $b = ۰$، در نتیجه اگر قرار دهیم $j = a_۱a_۲a_۳$ آنگاه $j^۳ = ۱$ و چون $g_۱g_۲g_۳ \notin K$، $j \neq ۱$. از طرف دیگر با استفاده از فرمول‌هائی که برای P، Q و R نوشتیم و با یک عملیات مستقیم می‌توان نشان داد که:

$$b = -ja_۱{}^۲a_۲(a_۱ - j)(a_۲ - j)(a_۳ - j)(P + Qj + Rj^۲)$$

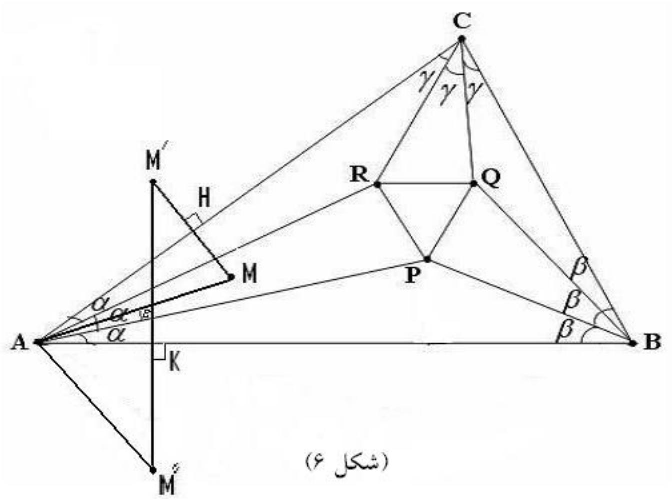
چون $a_i \neq j$ برای هر $i = ۱, ۲, ۳$ برقراراست پس $b = ۰$ اگر و تنها اگر $P + Qj + Rj^۲ = ۰$ و قضیه ثابت می‌شود.■

با استفاده از قضیه فوق قضیه مورلی ثابت می‌شود. در زیر فرض می‌کنیم $F = \mathbb{C} = \mathbb{R}^۲$ میدان اعداد مختلط است.

نتیجهٔ ۱. **(قضیه مورلی)** نقاط برخورد خطوط مجاور اضلاع تثلیت کنندهٔ سه زاویهٔ داخلی هر مثلث رأس‌های یک مثلث متساوی‌الاضلاع هستند.

برهان. مثلث ABC و تثلیث کننده‌های سه زاویه داخلی مثلث و نقاط برخورد خطوطمجاور

اضلاع، یعنی P، Q و R، را در شکل زیر در نظر می‌گیریم:



(شکل ۶)

قرار می‌دهیم $A = ۳\alpha$، $B = ۳\beta$ و $C = ۳\gamma$. فرض می‌کنیم $g_۱$ دوران به مرکز A و زاویه ۲α، $g_۲$ دوران به مرکز B و زاویه ۲β و $g_۳$ دوران به مرکز C و زاویه ۲γ باشد. تقارن‌های محوری با محورهای AB، BC و CA را به ترتیب با $S(AB)$، $S(BC)$ و $S(CA)$ نمایش می‌دهیم که هر یک عضوی مرتبه ۲ از $Af(\mathbb{C})$ می‌باشند.

اگر M نقطهٔ دلخواهی از صفحه باشد و قرینهٔ آن را نسبت به AC، M' و قرینهٔ M' نسبت به AB را M'' بنامیم آنگاه زاویهٔ $\widehat{MAM''}$ چنین محاسبه می‌گردد:

$$\widehat{MAM''} = \epsilon + \alpha + \widehat{KAM''} = \epsilon + \alpha + ۳\alpha + \widehat{HAM'}$$

$$= \epsilon + ۴\alpha + ۲\alpha - \epsilon = ۶\alpha$$

که در آن $\epsilon = \widehat{MAP}$.

بنابراین دیده می‌شود که M'' حاصل دوران M حول A با زاویه ۶α است و لذا $g_۱{}^۳ = S(AB)S(AC)$. به طور مشابه ثابت می‌شود $g_۲{}^۳ = S(AC)S(BC)$ و $g_۳{}^۳ = S(BC)S(AC)$.

اما چون هر تقارن محوری از مرتبهٔ ۲ است داریم

$$g_۱{}^۳ g_۲{}^۳ g_۳{}^۳ = S(AB)S(AC)S(AC)S(BC)S(BC)S(AC) = I$$

و لذا $g_۱{}^۳ g_۲{}^۳ g_۳{}^۳ = I$.

اما همانطور که توضیح دادیم و به راحتی هم دیده می‌شود P، Q و R به ترتیب نقاط ثابت تبدیلات

g_1g_2، g_2g_3 و g_3g_1 اند و در نتیجه طبق قضیه ۱ (ب) بالا حاصل می‌شود $P + Qj + Rj^2 = 0$. اکنون با استفاده از لم ۱ نتیجه می‌شود که مثلثِ PQR متساوی‌الاضلاع است و قضیه مورلی ثابت می‌گردد. ■

مراجع

[1] M. Berger, Geometry I, II, Springer-Verlag, 1987.

[2] A. Connes, A new proof of Morley's theorem, Les relations entre les mathematiques et la physique theorique, Inst. Hautes Etudes Sci., Bures-sur-yvette (1998), 43-46.

[3] A Connes, Symmetries, Newsletter European Mathematical Society, No. 54, Dec. 2004,pp.11-18.

[4] H. S. M. Coxeter and S. L. Greitzer, Geometry revisited, MAA, 1967.

[5] Educational times (New series) Vol.15 (1909), p.47

[6] C. Lubin, A proof of Morley's theorem, *Amer. Math. Monthly*, **62** (1955), 110-112.

[7] F. Morley, Extensions of Clifford's chain theorem, *Amer.J. Math.* vol.51,(1929), 465-472.

[8] D. O. Shklyarsky, N. N. Chenstov and Y. M. Yaglom, Selected problems and theorems of elementary mathematics, Vol. 2, Problem 97, Moscow, 1952.

[۹] حسین غیور، قضیهٔ مورلی، رشد آموزش ریاضی، سال اول، شماره ۲، تابستان ۱۳۶۳

[۱۰] حسین غیور، تعمیم قضیهٔ مورلی، رشد آموزش ریاضی، سال اول، شماره ۳، پاییز ۱۳۶۳

محمدرضا درفشه
دانشکده ریاضی، آمار و کامپیوتر، دانشگاه تهران
daraf@khayam.ut.ac.ir

فرهنگ و اندیشهٔ ریاضی
شماره ۴۱ (پاییز ۱۳۸۷) صص. ۵۳ تا ۵۶

بدرود بورباکی، تغییر دیدمان* در ریاضیات

یان استوارت

ترجمهٔ احسان ممتحن

توماس کوهن جایگاهش را از خاطر نشان ساختن این نکته کسب کرده است که علم دنباله‌ای پیوسته از تحولات عقلی نیست، بلکه تابعِ تغییر دیدمان است ــ انقلاب‌هایی که در آن‌ها تمام دیدگاه دگرگون می‌شود. واژهٔ خودمانی‌تر برای این فرآیند همان «مُد» است. اما مگر مد چیزی به جز تغییر دیدمان در زمینه‌ای معمولی از فعالیت‌های بشری است؟ اغلب گمان می‌رود ریاضیات از این تغییر دیدمان مصون است. هرچه که باشد، ریاضیات شامل حقایقی کامل است، که با منطقی انکارناپذیر اثبات شده‌اند. اما واقعاً این گونه است؟ خُب نه، نه واقعاً ــ اما اگر این جور هم بود، باز مانعی بر سر راه تغییر دیدمان نبود. تغییر دیدمان به منطق درونی علم مربوط نیست، بلکه با دیدگاه بیرونی آن مرتبط است. دوباره دیدن همان امور با نگاهی تازه و قرار دادن آنها در چهارچوبی نو است.

ریاضیات در تاریخ طولانی و پر نقش و نگارش تغییرهای دیدمان متعددی را تجربه کرده است، و هم‌اکنون در میانهٔ یکی دیگر از آنها است. دیدمان پیشین ریاضیات در دهه‌های پنجاه و شصت شکل گرفت، هنگامی که گروهی از ریاضیدانان با نام مستعار «بورباکی» بر آن شدند که تمام ریاضیات را بر بنیادی یکپارچه، کُلّی و بنابراین بسیار مجرد استوار سازند[۲] ایدهٔ آنها این بود که اصولی چنان کلی پایه‌ریزی کنند که از آنها تمام حالات خاصی که نسل‌های قبلی ریاضیدانان ابداع کرده‌اند قابل استنتاج باشد. برای مثال به جای مطالعهٔ هندسهٔ مسطحه، آن را به عنوان حالت خاصی از هندسهٔ فضاهای چند بعدی توصیف می‌کردند.

بورباکی آنقدر مزایا داشت که توانست خودش را همچون یک دیدمان تثبیت نماید. آنها دقت،

---

*) Paradigm Shifts 1) IAN STEWART

۲) بر طبق اظهار صریح آندره وی و دیگر اصحاب بورباکی نظیر پیر کارتیه، زمان پیدایش بورباکی به اوایل دههٔ چهل میلادی باز می‌گردد، نه دههٔ پنجاه یا شصت (م).

روشنی و اثر بخشی را در ریاضیات وارد کردند و حوزهٔ کاربرد آن را گسترش دادند. ریاضیات آنچنان غول پیکر و چهل تکه شده بود که کاری باید انجام می‌گرفت تا حقایقی را که آدمی باید می‌آموخت تقلیل دهد. بورباکی مثالی است از آنچه دانشمندان علوم کامپیوتر آن را اکنون فشرده‌سازی داده‌ها[1] می‌نامند. اگر روشی برای استنتاج حقایقی خاص از اصول کلی داده شده باشد، همهٔ آنچه که شخص نیاز دارد، به خاطر سپردن آن اصول است. یک کاستی اساسی وجود دارد، که نزد دانشمندان کامپیوتر هم شناخته شده است: داده‌های فشرده شده را، پیش از آن که بتوان با دقت بیان نمود، باید رمزگشائی کرد. ریاضیدانان می‌دانستند چگونه پیام‌های بورباکی را رمزگشائی کنند، اما مابقی عالم نمی‌دانست و این به سوء تفاهم اسف باری منجر شد، در پایان دههٔ شصت، میان بخش‌های ریاضی و فیزیک دیگر گفتگوئی در نمی‌گرفت. هر چند، در اواسط دههٔ هشتاد موانع در حال از میان برداشته شدن بود. در دههٔ نود، مدال فیلدز ـ معادل جایزهٔ نوبل در ریاضیات ـ به یک فیزیکدان داده شد، با وجود آن که بعضی از مهمترین ایده‌هایش اثبات ناشده باقی مانده بود.

چگونه چنین چیزی اتفاق افتاد؟ جابجائی‌های دیدمان زمانی ظهور می‌کنند که دیدمان قدیمی در جریان فکریِ رایج از اعتبار ساقط می‌شود. بورباکیسم به همین راحتی ناپدید نشد. قضایائی که با استفاده از روش‌های بورباکی پسند اثبات شده‌اند درست باقی می‌مانند حتی اگر شما دیگر به فلسفه‌ای که آنها را تولید کرده است معتقد نباشید. به این معنی، ریاضیات نو ریاضیات کهنه را زیر و رو نمی‌کند، و به همین دلیل است که عده‌ای از مردم فکر می‌کنند که تغییر دیدمان در ریاضیات رخ نمی‌دهد. اتفاقاً رخ می‌دهد و سقوط بورباکی هم نمونه‌ای از آن است. چیزی که پیش آمد آن بود که چشمهٔ قضایائی که می‌شد در کلیت مورد نظر بورباکی اثبات شوند به سرعت شروع به خشک شدن کرد. دستاورد بورباکی شامل تکنیک‌های فوق‌العاده‌ای بود، اما محدودیت‌های خودش را داشت ـ یکی از اصلی‌ترین محدودیت‌ها آن بود که بورباکی به نادیده گرفتن حالات خاص غیرعادی گرایش داشت، نتایج عجیب کم تعدادی که دربارهٔ یک مثال بودند. وضع قدری شبیه نظریهٔ عمومی خم‌ها است که چون دوایر تنها حالت خاصی از خم‌های به مراتب پیچیده‌ترند، اهمیت π را ارج نمی‌نهد. فیزیکدانی که برندهٔ مدال فیلدز شد یک آمریکائی است بنام اِد (مخفف ادوارد) ویتن، و بر روی نظریهٔ میدان کوانتومی و گره‌ها[2] کار می‌کند. کار بزرگ او کشف این بصیرت بود که این دو رشته باید به شکلی با یکدیگر مرتبط باشند. دست بر قضا کار ویتن نتیجهٔ مشاهدات دیگر برندهٔ مدال فیلدز، ریاضیدان اهل زلاند نو، وان جونز[1] بود. جونز در حالی که مشغول کار

1) data compression

۲) نظریهٔ گره‌ها شاخه‌ای از توپولوژی است. هر گره یک نشانندهٔ دایره در $\mathbb{R}^3$ است. ویتن در مقام یک متخصص نظریهٔ ریسمان، جهان را متشکل از ریسمان‌هائی از انرژی می‌داند که اغلب به شکل گره‌های خاصی در فضائی با ابعاد زیاد (گاه ۱۱ بعد) در حال نوسان و جنبش هستند. به همین علت درک ماهیت عمیق گره‌ها برای ادوارد ویتن مهم بوده و هست.(م)

3) Vaughan Jones

بر روی مسأله‌ای خاص در آنالیز بود با ریاضیات گره‌ها برخورد کرد. با تعقیب این مسأله، جونز روشی کاملاً جدید کشف کرد تا بتواند گره‌ها را از یکدیگر مشخص نماید. این روش به واقع دشوار بود: تا پیش از دههٔ بیست هیچ کس حتی نشان نداده بود که گره‌ها به معنی ریاضی کلمه وجود دارند، حتی کسی اثباتی منطقی ارائه نکرده بود که نشان دهد گره‌های ملوانی[1] و خِفتی[2] دو نوع گرهٔ متفاوت[3] هستند. گره‌ها بخشی از توپولوژی هستند و بورباکی به طرز پیروزمندانه‌ای بر توپولوژی فضاهای چندبعدی اشراف یافته و تقریباً همهٔ مسائل مهم در پنج یا بیشتر از پنج بعد را حل کرده بود. اما جونز توجه‌اش را به گره‌ها در فضای سه بعدی معطوف کرد و در آنجا چیزی فوق‌العاده یافت. شگفت‌انگیزترین جنبهٔ کشف جونز این است که در اساس کشف او می‌توانست پنجاه سال زودتر اتفاق افتد: کار جونز نتیجهٔ هیچکدام از نظریه‌های عمیق بورباکی نبود. تو گوئی در نوعی بُعد زمانیِ دیگر زندگی می‌کرد. مشاهدات جونز حلقهٔ جدیدی از ایده‌های اساسی را در سراسر ریاضیات گشود و کاربردهائی در زیست شناسی ملکولی DNA یافت. ویتن در حالی که تلاش می کرد بداند این ایده‌ها واقعاً چه می‌گویند، مشاهده کرد که خواستگاه حقیقی آنها نظریهٔ میدان کوانتومی است.

آواز طبل سقوط بورباکی همه جا به گوش می‌رسید. بورباکی نمای آراسته‌ای از ریاضیات را به نمایش می‌گذاشت که کاملاً با عملکرد واقعی ریاضیدانان مطابقت نداشت. ریاضیدانان پژوهشگر دربارهٔ حالات خاص فکر می‌کنند. آنها فضاهای چندبعدی را با فکر کردن دربارهٔ حالت $n = ۲$ مطالعه می‌کنند. اما ــ دست کم تا زمانی که دیدمان بورباکی غلبه داشت ــ آنها تنها شاهکارهای آراسته، نهائی و کلی خود را منتشر می‌کردند. نمونهٔ قدیمی‌ترِ استفاده از این شیوه کارل فریدریش گاوس بود، بزرگترین ریاضیدانی که تاکنون پا به عرصهٔ وجود نهاده است. یکی از معاصرانش او را همچون «روباهی که ردپاهایش را با دمش پاک می‌کند» توصیف کرده است. دیدمان کنونی از این قرار است: رد پای روباه مهم است. این که چگونه به ایده‌ای دست یابی به اندازهٔ خود ایده مهم است. بیان صریح‌تر این موضوع در «ریاضیات آزمایشگاهی» ممکن است، که تأکیدش بیشتر بر کامپیوتر یا کاغذ و مدادی است که برای بررسی مثال‌های خاص به کار می‌روند. بسیاری از پدیده‌های عام و مهم نظریهٔ آشوب دقیقاً به همین شکل پیدا شده‌اند: نخست آزمایش، سپس اثبات. ریاضیدانان آزمایشگر باور دارند که انتشارات علمی باید قصه را بدانگونه که رخ داده است بازگو کند، تا آن که پاسخ‌های روشنفکرانهٔ پس از واقعه را به نمایش بگذارد. خطاست اگر گمان بریم بورباکی یک اشتباه، یک اتلاف وقت بوده است. مسیر حرکت تاریخ مارپیچ است نه دایره: هنگامی که به دیدمان مشابه باز گردیم همان کار را در سطحی بالاتر انجام خواهیم داد. امروزه، حالات خاص از منظر بورباکی حالات خاص‌اند: اما از منظر ریاضیدانان قرن هیجدهم به طرز چشم‌گیری کلی و مجرد هستند. دیدمان جدید دستاوردهای مجرد بورباکی را همچون ابزارهائی جهت کشف

1) reef 2) granny

۳) گرهٔ ملوانی گره‌ایست که ملوانان در پائین یا بالای بادبان‌ها از آن استفاده می‌کنند و گرهٔ خفتی شبیه به گرهٔ طناب اعدام است. (م)

مثال‌های ملموس به خدمت می‌گیرد. قدرت همان ابزارها اوج باورنکردنی کنونی را میسر ساخته است. بناهای معظم بورباکی ویران نشده‌اند. ریاضیدانان در آنها زندگی می‌کنند، کار روزانه‌شان را در آنها انجام می‌دهند. اما به باور تعداد روزافزونی، آن کارها برای بنای ساختمان معظم دیگری به سبک بورباکی اختصاص ندارند.

البته، تا فرا رسیدن دیدمان بعدی، وقت آن است شروع به فهم ژرفای حقیقی عمومیّت حالات خاص کنیم.

مترجم: احسان ممتحن
گروه ریاضی ـ دانشگاه یاسوج
پست الکترونیک: momtahan_e@hotmail.com

فرهنگ و اندیشهٔ ریاضی
شماره ۴۱ (پاییز ۱۳۸۷) صص. ۵۷ تا ۶۸

نقد کتاب*

برایان بلنک

راه به سوی حقیقت: یک راهنمای کامل به قوانین جهان

نوشتهٔ راجر پنروز

نقاد: برایان بلنک[1]

ترجمه: سید محمد باقر کاشانی

راجر پنروز برای عنوان کتاب اخیرش، استعارهٔ راه به سوی حقیقت را برگزیده است که مکرراً در متون توصیفی فیزیک به کار می‌رود. تعجبی ندارد که این عبارت مورد پسند فیزیکدانان قرار گرفته است، زیرا این استعاره پیگیری سرراست مقصدهائی را پیشنهاد می‌کند که عبارت است از: درکی از همهٔ اصول اساسی که بر رفتار جهان حاکم است. شاید این خواسته برنامه‌ای بلند پروازانه به نظر آید. به ویژه با توجه به این گفتهٔ ای. ویگنر که «موفقیت بزرگ فیزیک به خاطر محدودیت هدف‌هایش می‌باشد.» از زمان این ارزیابی هوشیارانه، پیشرفت خیره کننده چنان چشم‌انداز فیزیک را تغییر داده است که پیشگامان زیادی از آن جسارت یافته‌اند که پیشنهاد کنند: درک کامل قوانین طبیعت در دسترس است.

چنان که پنروز بیان می‌کند، سفر (علمی) اکتشاف‌ها بیش از هزاران سال طول کشیده و بسیار مشکل بوده است. در آغاز این سفر، در حدود ۵۰۰ سال قبل از میلاد، هراکلیتوس مانع عمده را تشخیص داد: طبیعت عادت دارد خود را پنهان کند. صرف نظر از پیشرفت‌های ریاضی، گام‌های

---

*) Book Review, The Road to Reality: A Complete Guide to the Laws of the Universe, Reviewed by Brian Blank, Notices of the AMS, June/July 2006, 661-666

۱) برایان بلنک (Brian Blank) استاد ریاضی در دانشگاه واشنگتن است. آدرس الکترونیکی او عبارت است از: brian@math.wustl.edu

مهم اولیه در راه به سوی حقیقت در فاصلهٔ زمانی بین ۱۵۴۳، سالی که کوپرنیک نظریهٔ خورشید مرکزی خود را دربارهٔ حرکت سیارات منتشر کرد، و ۱۶۸۷ سال معرفی مکانیک نیوتنی برداشته شد. دورهٔ انقلاب علمی اول، منجر به آگاهی کارآمد از منظومهٔ شمسی، یک چارچوب اصلی برای مطالعهٔ دینامیک و یک فرمول‌بندی ریاضی از یکی از اندرکنش‌های طبیعت یعنی جاذبه شد. بسیاری از نشانه‌های پیشرفت در طول راه به سوی حقیقت را می‌توان در اولین انقلاب علمی اول تشخیص داد. از آنجا که طبیعت بسیار ماهرانه رازهایش را پنهان می‌کند، برای هر پیشرفت غالباً یک نظریهٔ جسورانه لازم است. مثلاً کوپرنیک، نه تنها قدرت حاکم بلکه حتی عقل سلیم هر ناظری را به مبارزه طلبید، وقتی مشاهده کرد که به خطا به جای احساس حرکت زمین، خورشید را در آسمان متحرک می‌بینند. لوازم دیگر پیشرفت ــ ریاضیات برای فرمول‌بندی و توسعهٔ نظریه، ابزار فیزیکی برای آزمایش کردن آن، نیز مؤلفه‌های اساسی انقلاب بودند. وسایل پیشرفتهٔ اندازه‌گیری، که به‌وسیلهٔ تیخو براهه طراحی شده بودند، کپلر را قادر ساخت تا مدارهای سیستم کوپرنیک را رد کند. حسابان نیوتن به او این امکان را داد که دینامیک گالیله را توسعه و قوانینی را که کپلر مشاهده کرده بود، توضیح دهد. راه به سوی حقیقت مسیر فعلی انقلاب را با پشت سرگذاردن تقریب‌های پی در پی برگزیده بود.

قرن نوزدهم شاهد انقلاب علمی دوم بود. بین سال‌های ۱۸۵۰ و ۱۸۶۵، مفاهیم بنیادی مانند انرژی و آنتروپی معرفی شد. ابتدا، بسیاری از دانشمندان بر علیه انرژی به عنوان یک تجرید ریاضی موضع گرفتند. ولی در پایان قرن، آن را ترجیحاً به عنوان مرجعی برای حقیقتی که نظریه‌های فیزیکی باید حول آن شکل بگیرد، جایگزین نیرو نمودند. شاخه‌های متعددی از فیزیک ــ ترمودینامیک، مکانیک آماری، نظریهٔ جنبشی گازها بر این اساس ظهور کردند. انقلاب دوم در فیزیک به آگاهی‌های ذیل منجر شد. آگاهی کارآمد از مکان منظومهٔ شمسی در کهکشان راه شیری، مفهوم میدان دافع، توصیف ریاضیِ اندر کنش بنیادی دوم، یعنی الکترومغناطیس، کشف یک ذرهٔ بنیادی (الکترون)، و بنا بر قانون دوم ترمودینامیک، صورت کاملاً جدیدی از واقعیت: جهت ترمودینامیکی زمان.

با وجود همهٔ موفقیت‌های انقلاب دوم، فیزیک در آغاز قرن بیستم با چالش‌های گوناگونی روبرو بود. بحث دربارهٔ ماهیت موجی در برابر ماهیت ذره‌ای نور، که در خلال انقلاب اول درگرفت، علیرغم توصیف ماکسول از نور به عنوان تشعشع الکترومغناطیسی، تا انقلاب دوم به طور قطعی سامان نیافت. همهٔ آزمایش‌ها برای کشف ماده‌ای که نور در خلال آن پراکنده می‌شود، فرضِ اِتِرِ درخشان، به شکست انجامید. قانون جاذبهٔ نیوتن به عنوان یک برداشت مفید از جاذبه باقی ماند، ولی این قانون نه مکانیسمی را که با آن جاذبه تحت تأثیر قرار می‌گیرد، توضیح می‌داد، نه به زمان اجازه می‌داد نقشی در عمل جاذبه ایفا کند. نظریه‌های جدید انقلاب دوم حتی پارادوکس‌های با دردسر بیشتری را به همراه داشت، مهمترین آن‌ها عبارت بود از حدس تشعشع جسم سیاه که انرژی به دلخواه بزرگ دارد. از دیدگاه تکاملی، به نظر می‌رسد درک ما از واقعیت پیشرفت کرده است. البته این پیشرفت نه با طی مسیر به سوی انتهای یک راه قدیمی بلکه با فراز و فرودهای پی در پی

بین وضعیت‌های دشوار هم وزن، حاصل شده است.

مورخین علم غالباً بیان می‌کنند که قرن بیستم شاهد دو انقلاب در فیزیک بود: نظریهٔ کوانتم و نسبیت عام. انقلاب اول فیزیک را از فاجعه تشعشع ماوراء بنفش جسم سیاه نجات داد و معمای ایجاد شده توسط ویژگی‌های نور را برطرف کرد. با از بین رفتن فرق بین موج و ذره، نظریهٔ کوانتم آگاهی‌های متضادی دربارهٔ ماهیت ماده و انرژی ارائه داد. انقلاب دوم، یعنی نسبیت عام، فضا و زمان را ترکیب کرد تا نظریه‌ای از جاذبه را ارائه دهد که بسیار عمیق‌تر از فرمولی است که در کتاب‌ها می‌نویسند. هر دو انقلاب درک ما از واقعیت فیزیکی را عمیقاً تغییر دادند.

قرن بیستم، در واقع، قرن پر باری برای فیزیک‌دانان بود؛ دو انقلاب علمی ممکن است هنوز انتظار آنان را برآورده نکرده باشد. فوتون، به‌عنوان ذرهٔ بنیادی دوم، در سال ۱۹۲۳ کشف شد. تا سال ۱۹۳۲ هر دو ذرهٔ پروتون و نوترون نیز کشف شد. این ذرات هسته‌ای فیزیکدانان را به کشف دو اندرکنش دیگر هدایت کردند: نیروهای قوی و ضعیف هسته‌ای. در خلال چند دهه، مجموعهٔ بزرگی از ذرات زیراتمی جمع‌آوری شد: پوزیترون‌ها و میون‌ها در دهۀ ۱۹۳۰، پیون‌ها و کِیون‌ها در دهۀ ۱۹۴۰، نوترینو (حدسیهٔ طولانی مدت پارلی) در دهۀ ۱۹۵۰ و خیلی ذرات دیگر. جدول در حال افزایش ذرات بیشتر گیج کننده می‌شد. پاولی ولفگنگ بلافاصله پس از ارائهٔ یک سخنرانی بحث‌انگیز در کلمبیا، پذیرفت که «این (جدول ذرات) یک نظریهٔ جسورانهٔ (تکامل نیافته) است.» نیلز بوهر، از (میان) حاضران در مخالفت گفت « متأسفانه، آن‌قدر هم جسورانه نیست!» در سال ۱۹۶۳ موریْ ژل ـ مَن و، مستقلاً جرج زویگ نظریه‌ای از ذرات بنیادی جزئاً باردار را (که ژل ـ مَن آن‌ها را کوارک می‌نامید) مطرح کردند که فقط به اندازهٔ کافی جسورانه است.

در یک دهه و نیم بعد، آنچه موسوم به مدل استاندارد ذرات بنیادی و اندرکنش‌هایشان بود، ظهور کرد. مدل استاندارد نظریه‌ای است که فیزیکدانان تجربی مکرراً آن را برای دقت استانداردها تأیید کرده‌اند. کشف ذرات همچنان ادامه دارد. قابل توجه‌ترین آن‌ها عبارتند از کوارک عالی در سال ۱۹۹۵ τ ـ نوترینو در سال ۲۰۰۰ ـ ولی آن‌ها مانند قرار گرفتن ذرات مصنوعی ساخت بشر در جدول تناوبی، در نظریهٔ ذرات بنیادی قرار می‌گیرند.

در خلال همان دورهٔ زمانی که فیزیکدانان انرژی بالا، کوچک‌ترین ذرات عالم را بررسی می‌کردند، منجمین و اختر فیزیکدانان درک ما را از بزرگترین اجرام واقعی از جمله جهان دگرگون کردند. تا سال ۱۹۲۳ منجمین وجود کهکشان‌های ماوراء راه‌شیری را تأیید کرده بودند. مشاهدهٔ اجرام آسمانی (دور) سرد همراه با نسبیت عام منجر به ظهور شاخهٔ جدیدی از فیزیک، کیهان‌شناسی شد، که ما را بیشتر دربارهٔ این که جهان چگونه به وجود آمد و چگونه معدوم می‌شود، آگاه می‌سازد. اگرچه فیزیک ذرات بنیادی و کیهان‌شناسی با اشیاء با اندازه‌های متضاد (از نظر کوچکی و بزرگی) سر و کار دارند، این دو حوزه به صورت پیچیده‌ای در هم تنیده‌اند. اطلاعات حاصل از مطالعهٔ فرایند زیراتمی مبنای درک فیزیک ستارگان و ترکیب ذرات سنگین در جهان است. برعکس، اجزاء اصلی رمزآلود جهان آزمایش‌های مهمی برای نظریهٔ ذرات فراهم می‌آورند.

سیاحت علمی خیلی سریعی که تاکنون انجام داده‌ایم بیان کنندهٔ بخش بسیار کوچکی است از آنچه که راه به سوی حقیقت در ۱۱۰۰ صفحه‌اش پوشش می‌دهد. هر کس صفحاتی از کتاب را مرور کند فوراً درمی‌یابد که مخاطبان تقریباً همسان بیش از قطور بودن کتاب، راه به سوی حقیقت را از سایر کتاب‌های توصیفی متمایز می‌کند. تا جایی که من می‌دانم پنروز تنها نویسنده‌ای است که مفاهیم پیچیده‌ای از فیزیک را به کمک مفاهیم پیچیدهٔ ریاضی معرفی می‌کند. حتی یک ریاضیدان با تجربه هم که به مطالعهٔ صفحه‌ای از کتاب که به شرح نمادهای تانسوری می‌پردازد، ممکن است از کتاب پنروز رویگردان شود. همچنان که مؤلف در مقدمه‌اش توضیح می‌دهد «آن چه من باید بگویم نمی‌تواند بدون کاربرد بخشی از نمادهای ریاضی و بهره‌برداری از مفاهیم اصیل ریاضی به صورت مستدلی ارائه شود.» این اعلان را پیروی از جملهٔ کپرنیک «ریاضیات برای ریاضیدانان نوشته می‌شود» نگیرید، نظر پنروز این است که ریاضیات را باید نه تنها برای ریاضیدانان بلکه برای هر کس که مایل است آن را یاد بگیرد، نوشت. برای انجام این کار، در مقدمه درس‌های جبرانی شروع می‌شود که معرفی اعداد گویا به عنوان رده‌های هم‌ارزی یک نمونه آن است.

شانزده فصل اول راه به سوی حقیقت اصولاً به ریاضیات مورد نیاز برای بیان فیزیک مدرن اختصاص یافته است. تا به حال به صفحهٔ ۳۸۳ رسیده‌ایم، به خوانندهٔ شجاع موضوع‌های زیادی در آنالیز، جبر و هندسه معرفی می‌شود. از این موضوع‌ها، نیاز به آنالیز در مقایسه با دیگر موضوعات نسبتاً کم است. حسابان، سری فوریه، ابرتابع‌ها، رویه‌های ریمانی و مقدار کافی از نظریهٔ توابع مختلط برای بیان قضیهٔ نگاشت ریمان. مطالب ضروری از جبر شامل کوآترنیون‌ها، جبرهای کلیفورد و گراسمانی، جبرخطی، گروه‌های تبدیل و مقدار کافی از نظریهٔ لی برای بحث گروه‌های کلاسیک و جبرهای لی و نمایش آن‌ها می‌باشد. موضوع‌های هندسهٔ دیفرانسیل سخت‌ترین قسمت است: انتقال موازی، ژئودزیک‌ها، خمیدگی، مشتق برونی، حسابان برخمینه‌ها، التصاق‌ها و کلاف‌های تاری. با آنچه تاکنون گفته شد، پنروز بخش قابل قبولی از دروس ریاضی دورهٔ کارشناسی را در بخش اول (یک سوم) کتابش گنجانده است. وقتی پنروز ادعا می‌کند « من به مطالب کتاب برای القای درک درست به خواننده خوشبینم » مایلیم سخن او را باور کنیم. به عنوان یک ارزش از عهدهٔ (خواندن) برآمدن برای خواننده‌ای که به محض ظاهر شدن یک فرمول ریاضی، مطالعه را متوقف می‌کند، پنروز پیشنهاد می‌کند «از همهٔ فرمول‌ها صرف نظر کنید و فقط جملات را بخوانید.» چنین توصیه‌ای مطمئناً به درستی خوشبینانه است زیرا به ندرت صفحه‌ای یافت می‌شود که در آن متن و ریاضی عمیقاً در هم نیامیخته باشند. خوانندگانی که از ریاضیات دوری می‌جویند با جستجوی آثار عالی زیادی که مخاطبان عمومی‌تری را هدف قرار داده‌اند و نیز جمع‌آوری درست این آثار چنان که زمینه‌های مشابه را پوشش دهند، موفق‌ترند. حتی کسانی که از نمادهای ریاضی فرار نمی‌کنند ممکن است متن‌های علمی به سبک برایان گرین یا استفان هاوکینگ را ترجیح دهند. ریاضیدانانی که به فیزیک علاقه‌مندند، ممکن است کتاب پنروز را مورد بررسی جدی‌تر قرار دهند. عده‌ای ممکن است فقط ۱۶ فصل اول را به طور کامل انتخاب و مطالعه کنند. دیگرانی که خواستار به روز کردن اطلاعات در چند موضوع هستند؛ مطالب پنروز را فراهم کنندهٔ پیش زمینهٔ مفیدی برای

فیزیک آینده می‌یابند.

بعد از ۳۸۲ صفحه مقدمات ریاضی (خارج از مسیر اصلی)، پنروز توجه خود را به انقلاب‌های علمی متعدد که فیزیک را در قرن بیستم متحول کردند، معطوف می‌دارد. قسمت دوم کتاب شامل ۳۵۲ صفحه به نسبیت عام، نظریهٔ کوانتم، فیزیک ذرات بنیادی و کیهان‌شناسی اختصاص دارد. گذر از ریاضیات به فیزیک تقریباً هموار است. بخشی از این همواری به خاطر این است که پنروز از روی تعصب دو موضوع (ریاضی و فیزیک) را از هم جدا نمی‌کند. اعداد کوانتمی در فصل هندسهٔ اعداد مختلط معرفی می‌شود، التصاق پیمانه‌ای در فصل کلاف‌های تاری ظاهر می‌شود، و در جهت دیگر، فضاهای هیلبرت، عملگرهای یکانی و همسازهای کروی در فصل نظریهٔ کوانتم یافت می‌شود. دلیل دیگر برای درهم‌آمیختن هموار ریاضیات و فیزیک این است که پنروز با زبان یک فیزیک ریاضی‌دان صحبت می‌کند: حتی وقتی موضوع‌هایی از فیزیک برای ما کاملاً آشنا نیست، روش و زبان ارائه آن آشناست. برای خواننده انتخابگر رفت و آمد بین ریاضی و فیزیک با مجموعه وسیعی از ارجاع‌های به قبل و بعد آسان می‌شود. برای خواننده‌ای که از بخشی از ریاضیات صرفنظر کرده است تا سریعاً به فیزیک برسد ارجاع‌های به مطالب قبلی مفید است. ارجاع‌ها به مطالب بعدی انگیزهٔ خوانندگان را برای تلاش در ریاضیات به ظاهر مجرد تقویت کند.

خوانندگانی که به فصل‌های نسبیت خاص، نسبیت عام و مکانیک کوانتمی رسیده‌اند شاهد یک بررسی عالی از این مطالب‌اند که مشحون از بینش‌های فیزیکی و زمینهٔ ریاضی است. به ویژه، چهار فصل پیاپی که با ذرهٔ کوانتمی شروع می‌شود و با کشف نظری پاد ذرات به وسیلهٔ پل دیراک به سرانجام می‌رسد به طور خاص روشنگر است. در پایان این فصل‌ها، پنروز نشان می‌دهد که چگونه ترکیب نسبیت خاص و نظریهٔ کوانتم منجر به حدس پاد ذرات می‌شود. پنروز به کمک نظریهٔ هامیلتونی نسبیتی یک ذرهٔ کوانتم با جرم سکون μ، معادلهٔ کلاین ــ گوردون برای نگاشت موج ψ، $(\Box + (\frac{\mu}{\hbar})^2)\psi = 0$ را به‌دست می‌آورد. او نشان می‌دهد که چگونه دیراک با کشف مجدد جبرهای کلیفورد، معادلهٔ کلاین ــ گوردون را به آنچه امروز معادلات دیراک و پاد ــ دیراک می‌نامیم تجزیه کرد. این پیش زمینهٔ ریاضی برای خواننده بینشی اصیل دربارهٔ حدس دیراک از وجود پوزیترون، پادذرهٔ الکترون فراهم می‌کند. پوزیترون فقط یک سال پس از حدس وجودش کشف شد.

نظریهٔ دیراک دربارهٔ الکترون یک نقطهٔ عزیمت طبیعی برای مدل استاندارد ذرات بنیادی و اندرکنش‌های آنهاست، موضوعی که به خوبی مورد توجه متون توصیفی قرار نگرفته است. یکی از مشکلات این است که مدل استاندارد پر از نکات فنی است، که بیشتر آن‌ها نامأنوس‌اند و شهودی نیستند. برای نوآموز حتی مشکل بیشتر همپوشانی عبارت‌هاست، چنان که جملهٔ زیر از راه به سوی حقیقت یک نمونهٔ آن است. «خانوادهٔ هَدْرون‌ها (نوعی از ذرات بنیادی) شامل فرمیون‌های موسوم به «باریون‌ها» و همچنین بوزون‌های معروف به «مِزون‌ها» می‌شود. به قول برایان گرین صدها ذرهٔ مختلف وجود دارد، که به شما سرگیجه می‌دهد یا چشمانتان را خیره می‌کند، و همگی آن‌ها هم از یک نظریهٔ پیچیدهٔ سازگاری ریاضی قابل بحث، پیروی می‌کنند. کوتاه سخن این که، مؤلفی که

می‌کوشد مدل استانداردی را برای مستمعین عمومی توضیح دهد با دام‌های زیادی مواجه می‌شود، پنروز نمی‌تواند همهٔ آن‌ها را به کناری گذارد. مثلاً بحث هدرون‌ها را در راه به سوی حقیقت در نظر گیرید. هدرون‌ها، بنابر تعریف عبارتند از ذراتی که از طریق نیروی هسته‌ای قوی، اندرکنش دارند. در صفحهٔ ۱۰۱، پنروز آن‌ها را چنین معرفی می‌کند: «... دیدگاه مدرن [این است] که ذرات با «اندرکنش قوی» موسوم به هدرون‌ها (پروتون‌ها، نوترون‌ها، π ـ نرون‌ها، و غیره) بنابر فرضی از کوارک تشکیل شده‌اند». نکات زیر یک گزیدهٔ کوتاه از کتاب است که تازه‌وارد (به مبحث) را دچار مشکل می‌کند:

- کاماها باید حدود وجه وصفی را مشخص کنند؛ همچنان که نقل شد، تعریف پنروز متضمن این است که مجموعهٔ هدرون‌ها شامل مجموعهٔ ذرات با اندرکنش قوی نیست. این ابهام کاملاً برطرف نمی‌شود (زیرا) ۵۰۰ صفحهٔ بعد، پنروز هنگام توصیف مجدد هدرون‌ها، می‌گوید آن‌ها زیرمجموعهٔ سره‌ای از ذرات با اندرکنش قوی محسوب می‌شوند.
- وارد کردن π ــ مزون‌ها در لیستی از ذرات شناخته شده برای تثبیت مفهوم هدرون‌ها تأثیر متضادی دارد: خلاصه نقل شده تنها مدخل اندیس‌دار برای π ــ مزون‌هاست. مشکل، هنگامی گسترش می‌یابد که پنروز بعداً، بدون توضیح اصطلاح مترادف، پیون‌ها را چهار بار هنگام صحبت از π ــ مزون‌ها (صفحه‌های ۴۳٦، ۴۳۷، ۴۹۴ و ٦۲۸) به کار می‌برد. تعریف مزون سرانجام در ص ٦۴٦ ظاهر می‌شود ولی برجسته نمی‌شود.
- ذرات زیراتمی بیان شده با عبارت «و غیره» تا بعد از صفحه ۵۰۰ آشکار نمی‌شوند.
- کاربرد عبارت «فرض می‌شود که تشکیل شده‌اند از» به جای عبارت واقعی‌تر «تشکیل شده انداز» گمراه کننده است. جمله به جای عبارت « حقیقت این است» با ایهام «دیدگاه مدرن این است» شروع شده است. آیا طفره رفتن بیش از حد ضروری است؟
- گزارهٔ نقل شده دربارهٔ هدرون‌ها در صفحهٔ ٦۴۵ به صورت « فرض می‌شود همهٔ هدرون‌ها از کوارک تشکیل شده‌اند»، تناقض آمیز است اگر توجه کنیم که پنروز بعداً بیان می‌کند که هر نِرون از یک کوارک و یک پادکوارک تشکیل شده است. به علاوه، گوی‌های ــ چسبان که بنابر اعتقاد در BNL و CERN و DESY کشف شده‌اند هدرون‌های بدون کوارک‌اند که فقط از گلوئُن‌ها تشکیل شده‌اند.

نکاتی که با بحث پیشین برجسته شد منحصر به همین‌ها نیست و به کتاب پنروز هم اختصاص ندارد: حتی بخش‌هایی از بسیاری از متون کاملاً پذیرفته شدهٔ مدل استاندارد مانند [۳] و [۱۰] آزار دهنده‌اند. آنچه پنروز را مفتخر به عنوان بهترین عمومی کنندهٔ فیزیک می‌کند، این است که او خواننده را به ساختار ریاضی بحث نزدیک‌تر می‌کند. با بحث التصاق‌های پیمانه‌ای و گروه‌های تقارن به صورت نابدیهی، او خوانندگانش را هدایت می‌کند که علت نامگذاری مدل استاندارد،

1) Big Bang

به نظریهٔ $Sv(۳) \times Sv(۲) \times v(۱)$ را درک کنند. با فصل حادثهٔ انفجار خلقت جهان[۱] و میراث ترمودینامیکی آن، پنروز قسمت دوم راه به سوی حقیقت را با بیان خلاصه‌ای از اطلاعات فعلی ما از کیهان‌شناسی به پایان می‌برد. به طور خلاصه، نظریهٔ بیگ بنگِ جهانِ در حال گسترش را آ. فریدمَن (۱۹۲۲) و مستقلاً جرج لِمَت (۱۹۲۷) که معادلهٔ اینشتین دربارهٔ جاذبه را بدون فرض اولیهٔ جهان ساکن حل کردند، ارائه داد.

کشف رادوین هابل (۱۹۲۹) از رکود کهکشان‌ها یک گواهی زودهنگام تجربی برای مدل فریدمن ـ لِمَت فراهم کرد. با وجود این، به نظر می‌رسد توضیح بیگ بنگ برای گسترش جهان فقط یک نظریهٔ جسورانهٔ دیگر باشد ــ خود لغت بیگ بنگ در دههٔ ۱۹۵۰ از یک ریشهٔ طعنه‌آمیز که توسط یک کیهان‌شناس به کار می‌رفت، سرچشمه گرفت. با آغاز سال ۱۹۶۴ هنگامی که تشعشع کیهانی نهانیِ مایکروویُو که به وسیلهٔ بیک بنگ حدس زده شده بود، کشف شد، گواهی‌های تجربی زیادی این نظریه را تأیید می‌کرد. بررسی سنتی کیهان‌شناسی این بحث کوتاه را به علت جزئیات بیشتر حجیم کرد. پنروز از سوی دیگر به جبران این که این مطالب را در فقط دو پاراگراف بیان می‌کند، سرانجام بیش از یک سوم فصل را به این مبحث اختصاص می‌دهد. شرح‌های استاندارد از بیگ بنگ خواننده را در جریان مراحل جهانِ سردِ زمان پلانک تا حال حاضر قرار می‌دهد. پنروز چنین نمی‌کند. مفهوم دوران سرد[۲] مختصراً در فصل مدل استاندارد ظاهر می‌شود ولی نقشش در تکامل آغازین جهان به روشنی بیان نمی‌شود. به این دلیل، نَه ترکیب هسته‌ای، نَه تشکیل ستاره‌ای راهی در راه به سوی حقیقت ندارد. بحث مختصری از تکامل اختری وجود دارد که فقط آفرینش حفره‌های سیاه را توصیف می‌کند. روی هم رفته، این فصل متمرکز بر ترمودینامیکِ چیستان‌های کیهان‌شناسیِ بیگ بنگ است که پنروز از دههٔ ۱۹۷۰ آن‌ها را مطرح و مطالعه کرده است. گرچه خوب است متخصصی داریم که اسرار جهان را که به مدت سه دهه ذهنش را به خود مشغول کرده است، بیان می‌کند، ضرر کم بها دادن به مباحث دیگر این است که اگر خوانندگان بخواهند این حقیقت را درک کنند که جهانی که هم‌اکنون در آن زندگی می‌کنیم از یک پلاسمای ذرات بنیادی به وجود آمده، باید به مراجع دیگر رجوع کنند.

چنان که گفته شد، راه به سوی حقیقت از ادغام سه کتاب در یک مجلد حاصل شده است. بخش سوم که تقریباً به اندازهٔ دو بخش اول طولانی است معطوف به مسیر پیش رو (فیزیک آینده) است. در اینجاست که پنروز کاملاً مطابق شهرت خود به عنوان بدعتگزار در بین فیزیکدانان رفتار می‌کند. مطابق ایده‌ای از کیهان‌شناسی که اکنون عموماً پذیرفته شده است، لحظه‌ای از زمان کمتر از $۱۰^{-۱۲}$ ثانیه «پس از بنگ» جهان شاهد یک دورهٔ «تورمی» با رشد نمایی بوده است که در آن اندازه‌اش با ضریب حداقل $۱۰^{۳۰}$ بزرگ می‌شد. این نظریه در سال ۱۹۷۹ توسط آ. گوث به عنوان جوابی به مسألهٔ مغناطیس یک قطبی که عدم سازگاری بین کیهان‌شناسی بیگ بنگ و نظریه‌های مشهور وحدت را مطرح می‌کرد، بیان شد. ابتدا، کیهان‌شناسی تورمی یک نظریهٔ جسورانهٔ دیگر

1) freezing out

به نظر می‌رسد. با وجود این، تورم مشکلات مهم چندی را از نظریهٔ سنتی بیگ بنگ برطرف کرد و ظرف مدت کوتاهی توجهی جدی به خود جلب نمود. تورم همچنین برشکاکانی که آن را یک نظریهٔ آزمایش‌ناپذیر که تسلیم مشاهده نمی‌شود می‌دانستند، غالب آمد. به علاوه، تورم با فرض صحت، به وجود میدان فرضی هیگز در زمان آغازین جهان که مدت‌هاست مورد جستجو می‌باشد، اشاره دارد. همچنان که ل. لِدِرمَن، یکی از پیشگامان جستجوگر ذرات اعلام کرده است، «اختر فیزیکدان‌ها یک شییء هیگز کشف کرده‌اند!». علیرغم این نتایج رضایت‌بخش (از نظریهٔ تورم)، وقتی پنروز تردید می‌کند و می‌گوید «دلایل قوی برای تشکیک بنیادی در کیهان‌شناسی تورمی وجود دارد»، یک مزاحم به نظر می‌رسد.

از جدال دربارهٔ تورم می‌توان پرهیز نمود، ولی این بحث با مطالب بخش پایانی کتاب پنروز سازگار است. تز او این است که پدیده‌های به اندازهٔ کافی توضیح داده نشده به قدری زیادند که نزدیک آخر راه به سوی حقیقت نیستیم. بیشترین هشدار به خاطر ماهیت موقت مدل استاندارد است. زیرا این نظریه ثابت‌هایی بنیادی از طبیعت را به عنوان پارامترهای درون داد نیاز دارد، این نظریه، حقیقت جهانی دیگر را درست مانند جهان ما توصیف می‌کند. این که اکنون به نظر می‌رسد مقدار ثابت‌های بنیادی دلخواه است و از پیش تعیین شده نیست یک نقص دانش فعلی ماست. مسألهٔ قابل توجه بیشتر عبارت است از تصادم بین نظریهٔ کوانتم و نسبیت عام در حوزه‌هایی که هر دو باید به کار روند. تا زمانی که یک نظریه تحقیق‌پذیر (قابل تصدیق و تکذیب) از جاذبهٔ کوانتمی ظاهر شود، باید قبول کنیم که از انتهای راه بسیار دور است. در واقع، اکنون انشعابی در این راه وجود دارد که باعث مجادله دربارهٔ ادامهٔ صحیح آن شده است. به هر یک از سه راه ممکن به جاذبهٔ کوانتمی فصلی اختصاص می‌دهد: نظریهٔ ریسمان، متغیرهای طوقه‌ای (LQG)، و نظریهٔ خودش یعنی نظریهٔ تویستور[1]. (این رویکردها در یک سطح مقدماتی‌تر توسط لی اِسْمُلین یک طرفدار LQG، نیز در مرجع سه راه به سوی جاذبهٔ کوانتمی [۸] مورد بحث قرار گرفته است. با توجه به این که کتاب اخیر متخصص نظریهٔ ریسمان برایان گرین [۱] بخشی با عنوان راه‌ها به سوی حقیقت دارد، نتیجه می‌گیریم که حداقل توافقی دربارهٔ استعاره‌ای که باید به کار رود وجود دارد.)

پنروز ممکن است در فصل کیهان‌شناسی تورمی از ادامهٔ بحث منصرف شده باشد، ولی در فصل نظریهٔ ریسمان با سلاح خود را تسلیم نظریهٔ رایج می‌کند. در اولین پاراگراف این فصل، پنروز می‌نویسد، «فیزیکدانان خیلی کمی پیش بینی می‌کنند که تغییرات بنیادی در چارچوب مکانیک کوانتمی به وجود آید. بلکه، آن‌ها دربارهٔ ایده‌های ظاهراً عجیب مانند نیاز به بعدهای اضافی در فضا ـ زمان بحث می‌کنند، یا دربارهٔ ذرات نقطه‌ای که با موجودات توسعه یافته مشهور به «ریسمان» باید جایگزین شوند. پنج بخش بعد، حمله بیانی افزایش می‌یابد: «از نظر بدگویان به شدت افراطی‌اش، [نظریهٔ ریسمان] تاکنون از نظر فیزیکی، مطلقاً موفقیتی نداشته است، و شانس کمی وجود دارد که در فیزیک آینده نقش مهمی ایفا کند.» گرچه ممکن است خواننده ظنین باشد که

1) twostor

پنروز این کلمات را از زبان بدگویان می‌گوید، حداقل چنین استنباط می‌کند که او «به جنبه‌های زیادی از برنامه‌های جاری نظریهٔ ریسمان نگاه مثبتی ندارد.» برخلاف اعتراض‌های نظری به نظریهٔ ریسمان که پنروز آن‌ها را مطرح می‌کند، بسیاری از فیزیکدانان از نظر اصولی معترفند: روش آن‌ها این است که نظریه‌هایی را که با تجربه آزمایش نمی‌شوند به عنوان مطلب فلسفی، مذهبی یا ریاضی کنار بگذارند. اِم.وِلْتْمَن، برندهٔ جایزهٔ نوبل در کتاب اخیرش دربارهٔ فیزیک ذرات [۱۰،ص ۳۰۸] با بیانی صریح مانند پنروز قضاوتش را برای حذف نظریهٔ ریسمان و ابر تقارن اظهار می‌دارد: «حقیقت این است که این کتاب دربارهٔ فیزیک است و این امر متضمن این است که ایده‌های نظری که مورد بحث قرار می‌گیرند باید با حقایق تجربی پشتیبانی شوند. نَه ابر تقارن، نَه نظریه ریسمان این شرط را ارضا نمی‌کنند. آن‌ها آفریدهٔ تفکر نظری‌اند. به قول پائولی: آن‌ها حتی غلط نیستند.»

تفسیر آفرینش جهان و فرجام نهایی آن ناخواسته فکر نویسنده را به فلسفه، الهیات یا دیگر روزنه‌های فکری متوجه می‌کند. پنروز نَه به این می‌اندیشد که چرا در مقابل نیستی چیزی وجود دارد، نه در آن چه روزنامه نگار علمی تیموتی فریس آن را مَتاع ــ خداوندی می‌نامد، درگیر می‌شود. در عوض، راه به سوی حقیقت با فصلی خاتمه می‌یابد که در آن پنروز مجذوب و شیفتهٔ زیبایی و معجزه‌ها می‌شود. این‌ها عبارتند از فیزیک ریاضی گونه هدایت شده، نقش ابطال پذیری در نظریهٔ علمی، و نقش سلیقهٔ غالب در نظریهٔ فیزیکی. در خصوص آخرین موضوع، پنروز فَراری را مطرح می‌کند که در طول مسیر نظریهٔ ریسمان به جاذبهٔ کوانتمی در حال وقوع است. او نگران است که جریان فعلی تحقیقات در حال منحرف کردن تعداد فزاینده از نظریه‌دان‌ها(ی فیزیک) به مسیری است که به ظن او بن‌بست است. پنروز همچنین نگران پرگویانی است که می‌خواهند دورنما(ی فیزیک) را ترسیم کند. اگر شما قبلاً نظریهٔ ریسمان را مطالعه کرده باشید محتملاً با این تکیه کلام آشنایید. «نظریهٔ ریسمان تنها بازی در شهر است.» پنروز نقل می‌کند که یک کتاب درسی مشهور نظریهٔ ریسمان با بیانی مبارزه‌طلبانه اعلام می‌کند، «هیچ گزینهٔ دیگری وجود ندارد ... همهٔ ایده‌های خوب بخشی از نظریهٔ ریسمان است» مشابهاً، نویسنده یک کتاب جدید دربارهٔ نظریه ریسمان می‌گوید [۹، ص ۳۵۷]، «هر چند مایلم با توضیح جنبه‌های مخالف بین ایده‌ها، تعادل ایجاد کنم، به سادگی نمی‌توانم آن جنبه‌ها را بیابم» به خاطر فراوانی طرفداران، پنروز نمی‌تواند کاری انجام دهد جز یادآوری این که «(در برخورد) با ایده‌هائی که مانند ایده‌هایی در جاذبهٔ کوانتمی از تأیید یا رد تجربی دورند، باید محتاط باشیم تا شهرت یک رویکرد را به عنوان نشانهٔ صحت آن به حساب نیاوریم.» به علت ارائهٔ نظریه‌های متضاد توسط پنروز، ریاضیدانی که کتاب را مطالعه می‌کند نمی‌تواند تصمیم بگیرد کدام یک از این نظریه‌ها درست است، بنابراین به این موضع افراطی تمایل می‌یابد که این مطالب، به جدالی فلسفی مربوط‌اند.

تقریباً قطعی است که یک کتاب خیلی قطور باعث ناخوشایندی شود. ولی به ویژه به نظر من، متخصصی که فهرست اصطلاحات و اعلام کتاب پنروز را تهیه کرده است به دلایل زیر به خوبی از عهدهٔ این کار برنیامده است. الکترون با جرم سکون نوترینو $m(\nu_e)$ یک نمونهٔ خوب است. تعیین مقدار این پارامتر، و به‌ویژه، مشخص کردن این که این پارامتر ناصفر است در حال حاضر مورد علاقه

زیادی است. پنروز جرم نوترینو را در صفحه‌های ٦٣٦ و ٦٣٧ معرفی می‌کند، سپس کران بالایی برای $m(\nu_e)$ در صفحه ٨٧٢ ارائه می‌دهد. با وجود این، این بحث دوم هیچ مدخل مشخص ندارد و بحث اول زیر عنوان به اشتباه هجی شدهٔ nutrino مشخص شده است. نکتهٔ منفی دیگر عبارت است از معرفی اتفاقی فیزیک‌دانان نظریه‌پرداز. بسیاری از نام‌های آنان به حروف اولشان تقلیل یافته است، و بسیاری موارد حتی به این صورت هم ثبت نشده است: جیمز کرُونین، وال فیچ، جان کلیو وارد و جرج زویگ به ندرت در مدخل، یا با حرف اول، یا با نام اول یافت می‌شوند. فیزیکدانان دیگری کاملاً فراموش شده‌اند. بنابراین، از دستگاه اشترن ــ گرلاچ مطلعیم، ولی با نام دو پیشگام مهم نظریهٔ کوانتم والتر گرلاچ و اتو اشترن مواجه نمی‌شویم. در واقع، غالباً تصادفی است که نام یک فیزیکدان برجسته ذکر می‌شود. مثلاً ای. ویگنر دو مدخل دارد، ولی نه در فصلی که شامل ایده‌های فراوانی است که زمانی به ویگنریسم مشهور بود.

اگرچه موارد اعتراضی مذکور بجاست، اما در برابر تلاش شگفت‌انگیز پنروز اندک به نظر می‌رسند. خطاهای نسبتاً کم ذکر شده گواه بی‌دقتی کلی نیست. راه به سوی حقیقت با توجه به حجم کتاب، وسعت نظر در ریاضی و فیزیک هر دو و هشت سال دوران تألیفش، فوق‌العاده دقیق و سازگار به نظر می‌رسد. من فکر می‌کنم هر اعتراضی بجا (دربارهٔ مطالب کتاب) باید با توجه به عدم تعادل بین توصیف ریاضی قوانین فیزیکی و ارائه پشتیبانی‌های تجربی از آن‌ها صورت گیرد. اندیشه‌های نظری بخشی از ریاضی است ولی هنگامی که نوبت به تمایز بین نظریه‌های جسورانه بیانگر حقایق فیزیکی و نظریه‌های صرفاً جسورانه می‌رسد، برای حکم کردن باید به حقایق تجربی تکیه کنیم. حتی دیراک، کسی که معادلهٔ پادالکترون را معرفی نمود، برای بیان حدس پادماده تردید داشت. (چون او بعداً اعلام کرد، «معادله از من زیرک‌تر بود.») شاید بتوان تمرکز پنروز بر نظریه را این‌گونه آشکار نمود که وی عنوان شتاب دهندهٔ بزرگ سرن را که قرار بود در ٢٠٠٧ عملیاتش آغاز شود، از فصلی با عنوان راه حقیقت در کجا قرار دارد؟ حذف کرده است. برای کسب دیدگاه‌های تجربی فیزیک، خوانندگان می‌توانند به عنوان مکمل (کتاب) راه به سوی حقیقت کتاب‌های عالی لی.لدرمن و دُن لینکلن هر دو از آزمایشگاه فرمی ([۴] و [۵]) را مطالعه کنند. به ویژه مرجع [۲] کتاب ارزشمندی است که حاوی مقالاتی از فیزیک‌دانان تجربی و نظری است که در شکل‌گیری مدل استاندارد سهیم بوده‌اند.

(کتاب) راه به سوی حقیقت در بریتانیا چاپ شده و قبل از انتشار چاپ آمریکایی آن، توسط منتقدان مورد بحث قرار گرفته است. قبل از این که کتاب به دستم برسد، نقدهای متعددی از آن را خوانده بودم که با توصیف ضعیف خود آن را بد جلوه داده بودند. نمونه‌ای از چنین ارزیابی‌هایی ارزیابی زیر از فرنک ویلچک، برندهٔ جایزهٔ نوبل است [۱۱]. او می‌گوید: «مطالب قابل تحسین و مفید زیادی در این کتاب وجود دارد، ولی با رعایت بالاترین استانداردها، راه به سوی حقیقت عمیقاً معیوب است.» با اطلاع از چنین انتقادی، من نقدم را با نگرانی شروع کردم. ولی با پیشرفت در خلال چند صد صفحهٔ اول دریافتم که راه به سوی حقیقت با طرح خوبی نوشته شده است و همهٔ

داده‌های غلطم تسلیم تیزبینی‌های ارزشمند پنروز شد و این که او تکامل بنیادی فیزیک قرن بیستم را به خوبی بیان کرده است. ویلچک در نقدش می‌گوید «اشتباه‌های لپی جدی» توجه را به سه مطلب جلب می‌کند. این مسائل که فقط به خاطر پیچیدگی‌های فیزیک قابل درک است، از دید خوانندگان خاصی قابل طرح است.

اگر پنروز عامهٔ خوانندگانش را به درک مفاهیم مجادله‌آمیز نایل کرده باشد در مقایسه با اشتباه‌های تصادفی‌اش خدمت ماندگاری انجام داده است.

برای ریاضیدانان با علاقهٔ عمومی به فیزیک، کتاب پنروز به عنوان کتاب خودخوان توصیه می‌شود. برای ریاضیدانان دیگر مطالعهٔ اجمالی راه به سوی حقیقت با این هدف که ببینند ساختارهای ریاضی تا چه اندازه بر نظریه‌های فیزیکی اثر می‌گذارد، مفید است. کتاب‌های رقیب زیادی وجود دارد که در پی توضیح جایگاه هنر در فیزیک بنیادی هستند. اگر کتاب پنروز را با هر یک از کتاب‌های جدید (مثلاً با [٦]، [٧]، [٩]) مقایسه کنید، نظر این نقاد را مبنی بر این که **راه به سوی حقیقت** از بالاترین استانداردها برخورد است، درک می‌کنید. زیرا واقعاً این کتاب منحصر به فرد است. و این درک، آخرین توصیه مرا آسان می‌کند. برای هر کس که می‌خواهد فیزیک را به روز و در سطح استاندارد و پیشرفته بداند، راه به سوی حقیقت تنها کتاب موجود است.

تشکر مترجم: لازم است از نقاد کتاب، برایان بلنک که با پاسخ دادن به وسیلهٔ email به بعضی سؤال‌های اینجانب مرا در فهم بخش‌هایی از نقد و در نتیجه ترجمهٔ درست آن‌ها کمک کرده است و همچنین از ویراستاران فرهنگ و اندیشهٔ ریاضی، سپاسگزاری نمایم.

مراجع

[1] BRIAN GREENE, *The Fabric of the Cosmos*, Knopf, 2004.

[2] LILLIAN HODDESON, LAURIE BROWN, MICHAEL RIORDAN, and MAX DRESDEN, *The Rise of the Standard Model: Particle Physics in the 1960s and 1970s*, Cambridge University Press, 1997.

[3] GORDON KANE, *The Particle Garden*, Basic Books, 1995.

[4] LEON LEDERMAN With DICK TERESI, *The God Particle: If the Universe Is the Answer, What Is the Question?*, Houghton Mifflin, 1993.

[5] DON LINCOLN, *Understanding the Universe from Quarks to the Cosmos*, World Scientific Publishing Company, 2004

[6] ROBERT OERTER, *The Theory of Almost Everything: The Standard Model, the Unsung Triumph of Modern Physics*. Pi Press, 2006.

[7] LISA RANDALL, *Warped Passages: Unraveling the Mysteries of the Universe's Hidden Dimensions*, Harper Collins Publishing, 2005.

[8] LEE SMOLIN, *Three Roads to Quantum Gravity*, Basic Books, 2001.

[9] LEONARD SUSSKIND, *The Cosmic Landscape: String Theory and the Illusion of Intelligent Design*, Little Brown, 2006.

[10] MARTINUS VELTMAN, *Facts and Mysteries in Elementary Particle Physics,* World Scientific Publishing Company, 2003.

[11] FRANK WILCZEK, Review of *The Road to Reality: A Complete Guide to the Laws of the Universe*, by Roger Penrose (Jonathan Cape, Londan, 2004), *Science* 307 (11 February 2005), no. 5711, pp. 852 - 853.

مترجم: سید محمدباقر کاشانی
دانشگاه تربیت مدرس ـ دانشکده علوم‌پایه
kashanim@modares.ac.ir