

بسم الله الرحمن الرحيم



دانشگاه حکیم بسزوری

دانشکده ریاضی و علوم کامپیوتر

پایان نامه برای دریافت درجه کارشناسی ارشد در رشته علوم تصمیم و مهندسی دانش
گرایش

روشی مبتنی بر گرادیان برای افراز بندی طیفی گراف

استادان راهنما

دکتر محمود امین طوسی و دکتر مهدی زعفرانیه

پژوهشگر:

مهدی نعمتی

بهمن ۱۳۹۶



دانشگاه آزاد اسلامی

باسمه تعالی

فرم ارزشیابی و صورتجلسه دفاع از پایان نامه کارشناسی ارشد

فرم ۱۱۳-ت

جلسه دفاع از پایان نامه آقای /خانم مهدی نعمتی دانشجوی رشته علوم تصمیم و مهندسی دانش گرایش به شماره دانشجویی ۹۴۱۳۱۳۷۰۳۹ با عنوان:

روشی مبتنی بر گرادیان برای افزایش افرابندی طیفی گراف

در مورخه در دانشکده ریاضی و علوم کامپیوتر تشکیل و توسط هیات داوران مورد ارزشیابی قرار گرفت و نمره برابر درجه برای آن تعیین گردید .
به این ترتیب از این تاریخ آقای/ خانم مهدی نعمتی به عنوان کارشناس ارشد در رشته مذکور شناخته می شود .

مورد ارزشیابی	موارد	حداکثر نمره	نمره کسب شده
۱- کیفیت نگارش	رعایت اصول نگارش انسجام در تنظیم بخشهای مختلف، کیفیت تصاویر، جداول و اشکال، تنظیم فهرست ها، منابع و ماخذ.	۴	
۲- کیفیت علمی	بررسی تاریخچه و سابقه تجربی و نظری موضوع انسجام منطقی در بخش های مختلف پایان نامه، ابتکار و نوآوری، اهمیت و ارزش علمی پایان نامه، استفاده از منابع معتبر و جدید، کیفیت تجزیه و تحلیل یافته ها و نتیجه گیری، روشن بودن روش کار، هدف ها و فرضیه های تحقیق، جدید بودن روش تحقیق	۱۰	
۳- کیفیت ارائه در جلسه دفاع	تسلط بر موضوع و بیان واضح و تفهیم آن، توانایی در پاسخگویی به سوالات مطرح شده در جلسه، رعایت زمان ارائه، روش ارائه	۴	
۴- ارزشیابی گزارشات	گزارش های دوره ای پیشرفت کار (حداقل ۴ مورد)	۱	
۵- خروجی پایان نامه	مقاله مستخرج از پایان نامه: این نمره به صورت زیر اختصاص می یابد (۱) چکیده کنفرانسی هر مورد ۰/۲۵ نمره تا سقف ۰/۵ نمره (۲) مقاله کامل در مجموع مقالات همایشهای معتبر یا مقاله در مجلات علمی-ترویجی معتبر پذیرفته شده یا چاپ شده هر مورد ۰/۵ نمره تا سقف ۱ نمره (۳) مقاله پذیرفته شده یا چاپ شده در مجلات علمی پژوهشی معتبر ۱ نمره (۴) مقاله ارسال شده به مجلات علمی پژوهشی معتبر هر مورد ۰/۲۵ نمره تا سقف ۰/۵ نمره (۵) دستگاه ساخته شده دارای گواهی ثبت اختراع یا به سفارش سازمان ها تا سقف ۱ نمره (۶) دستگاه ساخته شده کاربردی که به تأیید رئیس دانشکده رسیده باشد تا سقف ۰/۵ نمره	۱	
جمع			

درجه معادل کسب شده: (از ۲۰ تا عالی) از ۱۸ تا ۱۸/۹۹ بسیار خوب از ۱۶ تا ۱۷/۹۹ خوب از ۱۴ تا ۱۵/۹۹ قابل قبول کمتر از ۱۴ غیر قابل قبول

مشخصات هیات دوران

ردیف	نام و نام خانوادگی	سمت	مرتبۀ علمی	محل کار	امضا
۱	دکتر محمود امین طوسی	استاد راهنما	استادیار	دانشگاه حکیم سبزواری	
۲	دکتر مهدی زعفرانی	استاد راهنما	استادیار	دانشگاه حکیم سبزواری	
۳	دکتر یاسر علیزاده ثانی	استاد داور	استادیار	دانشگاه حکیم سبزواری	
۴	دکتر	نماینده تحصیلات تکمیلی	استادیار	دانشگاه حکیم سبزواری	

امضا

رئیس دانشکده

امضا

مدیر گروه



سوگند نامه دانش آموختگان دانشگاه حکیم سبزواری

به نام خداوند جان و خرد کزین برتر اندیشه بر نگذرد

اینک که به خواست آفریدگار پاک، کوشش خویش و بهره گیری از دانش استادان و سرمایه‌های مادی و معنوی این مرز و بوم، توشه‌ای از دانش و خرد گردآورده‌ام، در پیشگاه خداوند بزرگ سوگند یاد می‌کنم که در به کارگیری دانش خویش، همواره بر راه راست و درست گام بردارم. خداوند بزرگ، شما شاهدان، دانشجویان و دیگر حاضران را به عنوان داورانی امین گواه می‌گیرم که از همه دانش و توان خود برای گسترش مرزهای دانش بهره‌گیرم و از هیچ کوششی برای تبدیل جهان به جایی بهتر برای زیستن، دریغ نورزم. پیمان می‌بندم که همواره کرامت انسانی را در نظر داشته باشم و هموعان خود را در هر زمان و مکان تا سر حد امکان یاری دهم. سوگند می‌خورم که در به کارگیری دانش خویش به کاری که باره و رسم انسانی، آیین پرهیزگاری، شرافت و اصول اخلاقی برخاسته از ادیان بزرگ الهی، به ویژه دین مبین اسلام، مابینت دارد دست نیازم. همچنین در سایه اصول جهان شمول انسانی و اسلامی، پیمان می‌بندم از هیچ کوششی برای آبادانی و سرافرازی میهن و هم میهنانم فروگذاری نکنم و خداوند بزرگ را به یاری طلبم تا همواره در پیشگاه او و در برابر وجدان بیدار خویش و ملت سرافراز، بر این پیمان تا ابد استوار بمانم.

نام و نام خانوادگی: مهدی نعمتی

تاریخ و امضا:

تأییدی صحت و اصالت نتایج

باسمه تعالی

اینجانب مهدی نعمتی به شماره دانشجویی ۹۴۱۳۱۳۷۰۳۹ دانشجوی رشته علوم تصمیم و مهندسی دانش مقطع تحصیلی کارشناسی ارشد تأیید می‌نمایم که کلیه نتایج این پایان‌نامه حاصل کار اینجانب و بدون هرگونه دخل و تصرف است و موارد نسخه برداری شده از آثار دیگران را با ذکر کامل مشخصات منبع ذکر کرده‌ام. در صورت اثبات خلاف مندرجات فوق، به تشخیص دانشگاه مطابق با ضوابط و مقررات حاکم (قانون حمایت از حقوق مؤلفان و مصنفان و قانون ترجمه و تکثیر کتب و نشریات و آثار صوتی، ضوابط و مقررات آموزشی، پژوهشی و انضباطی ...) با اینجانب رفتار خواهد شد و حق هرگونه اعتراض در خصوص احقاق حقوق مکتسب و تشخیص و تعیین تخلف و مجازات را از خویش سلب می‌نمایم. در ضمن، مسئولیت هرگونه پاسخگویی به اشخاص اعم از حقیقی و حقوقی و مراجع ذی صلاح (اعم از اداری و قضایی) به عهده ی اینجانب خواهد بود و دانشگاه هیچ گونه مسئولیتی در این خصوص نخواهد داشت.

نام و نام خانوادگی: مهدی نعمتی

تاریخ و امضا:

مجوز بهره برداری از پایان نامه

بهره برداری از این پایان نامه در چهارچوب مقررات کتابخانه و با توجه به محدودیتی که توسط استاد راهنما به شرح زیر

تعیین می شود، بلامانع است:

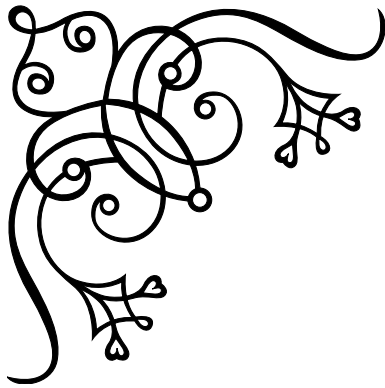
- بهره برداری از این پایان نامه برای همگان بلامانع است.
- بهره برداری از این پایان نامه با اخذ مجوز از استاد راهنما، بلامانع است.
- بهره برداری از این پایان نامه تا تاریخ ممنوع است.

استادان راهنما: دکتر محمود امین طوسی

دکتر مهدی زعفرانیه

تاریخ و امضا:

تقدیم به:



پدر و مادرم



سپاس خداوندگار حکیم را که با لطف بی کران خود، آدمی را زیور عقل آراست. در آغاز وظیفه خود می دانم از زحمات بی دریغ استادان راهنمای خود، جناب آقای دکتر محمود امین طوسی و جناب آقای دکتر مهدی زعفرانی، صمیمانه تشکر و قدردانی کنم که قطعاً بدون راهنمایی های ارزنده ایشان، این مجموعه به انجام نمی رسید.

همچنین لازم می دانم از گروه پارسی لاتک در پاسخگویی به مشکلات کاربران کمال قدردانی را داشته باشم. در پایان، بوسه می زنم بر دستان خداوندگاران مهر و مهربانی، پدر و مادر عزیزم و بعد از خدا، ستایش می کنم وجود مقدس شان را و تشکر می کنم از خانواده عزیزم به پاس عاطفه سرشار و گرمای امیدبخش وجودشان، که بهترین پشتیبان من بودند.

مهدی نعمتی

بهمن ۱۳۹۶

فهرست مطالب

ج	فهرست جداول
د	فهرست تصاویر
۱	چکیده
۲	پیش‌گفتار
۳	فصل ۱: تعاریف اولیه و پیش‌نیازها
۴	۱-۱ تاریخچه‌ای از مسأله خوشه‌بندی
۴	۱-۱-۱ کاربرد خوشه‌بندی
۵	۲-۱ گراف
۷	۳-۱ ماتریس
۱۲	۴-۱ نُرم
۱۳	۱-۴-۱ انواع نرم
۱۳	۵-۱ نرم‌های ماتریسی
۱۴	۶-۱ تجزیه مقدار تکین
۱۶	۷-۱ تقریب مرتبه پایین
۱۸	۸-۱ ماتریس لاپلاسیان گراف
۲۱	۹-۱ افراز گراف
۲۳	۱۰-۱ الگوریتم k-means برای خوشه‌بندی داده‌ها
۲۵	فصل ۲: روش‌های افرازبندی طیفی گراف
۲۵	۱-۲ افرازبندی طیفی گراف
۲۷	۲-۲ مثال
۳۱	۳-۲ برش k تایی همزمان با چند بردار ویژه

- ۴-۲ فرمول‌بندی روش طیفی برای بکارگیری روش‌های بهینه‌سازی ۳۲
- ۵-۲ مسئله بهینه‌سازی نامقید، معادل با مسئله (۷-۲) ۳۴
- ۶-۲ روش گرادیان نزولی ۳۵
- ۱-۶-۲ معیار همگرایی ۳۵
- ۷-۲ روش گرادیان مزدوج ۳۵
- ۱-۷-۲ الگوریتم فلچر-ریویس ۳۷
- ۸-۲ به‌کارگیری روش گرادیان مزدوج برای افزایش طیفی گراف ۴۲
- ۱-۸-۲ روش مینیمم‌سازی گرادیان مزدوج برای افزایش طیفی گراف ۴۳
- ۹-۲ روش نیستروم برای تخمین مقدار ویژه و بردار ویژه ۴۴
- ۱۰-۲ روش نمونه‌گیری افزایشی بر پایه احتمال ۴۷
- ۱۱-۲ الگوریتم افزایشی طیفی نیستروم مبتنی بر نمونه‌گیری احتمالی افزایشی ۴۹

فصل ۳: نتایج آزمایشات و روش‌های پیشنهادی ۵۲

- ۱-۳ استفاده از الگوریتم ژنتیک در گرادیان مزدوج ۵۲
- ۲-۳ پیاده‌سازی روش نیستروم بر روی مجموعه داده‌های آیریس و سرطان سینه ۵۵
- ۳-۳ موازی‌سازی الگوریتم بروزرسانی احتمال ۵۶

فهرست منابع ۵۷

پیوست آ: اثبات نمونه‌گیری افزایشی مبتنی بر احتمال ۵۹

پیوست ب: کد برنامه‌های روش گرادیان مزدوج و روش نیستروم ۶۵

واژه‌نامه فارسی به انگلیسی ۷۰

واژه‌نامه انگلیسی به فارسی ۷۲

فهرست جداول

۳-۳	زمان اجرا الگوریتم گرادیان مزدوج و الگوریتم ژنتیک	۵۴
۲-۳	تعداد یال‌های برشی گراف توسط الگوریتم گرادیان مزدوج و ژنتیک	۵۴
۱-۳	هزینه تابع هدف الگوریتم‌های گرادیان مزدوج و ژنتیک	۵۴
۶-۳	زمان اجرای الگوریتم نیستروم برای آپریس	۵۶
۵-۳	بهترین درصد خطای الگوریتم نیستروم برای آپریس	۵۶
۴-۳	درصد خطای الگوریتم نیستروم برای آپریس	۵۶
۹-۳	زمان اجرای الگوریتم نیستروم برای سرطان سینه	۵۶
۸-۳	بهترین درصد خطای الگوریتم نیستروم برای سرطان سینه	۵۶
۷-۳	درصد خطای الگوریتم نیستروم برای سرطان سینه	۵۶

فهرست تصاویر

۶	یک $v_1 v_6$ برش با $S = \{v_1, v_2, v_4\}$ و $\bar{S} = \{v_3, v_5, v_6\}$	۱-۱
۲۰	گراف با ۵ رأس	۲-۱
۲۲	یک گراف افزاز شده	۳-۱
۲۴	مراحل روش k-means برای خوشه‌بندی داده‌ها	۴-۱
۲۷	گراف با ۸ رأس	۱-۲
۳۰	افزاز حاصل از روش طیفی	۲-۲
	مثال‌های از عملکرد درست و نادرست الگوریتم گرادیان مزدوج در افزاز گراف. خط‌چین‌ها معرف	۱-۳
۵۳	لبه‌های مورد برش در گراف هستند.	



دانشگاه گیلان

فرم چکیده ی پایان نامه ی دوره ی تحصیلات تکمیلی

مدیریت تحصیلات تکمیلی

نام خانوادگی دانشجو: نعمتی	نام: مهدی	ش. دانشجویی: ۹۴۱۳۱۳۷۰۳۹
استادان راهنما: دکتر محمود امین طوسی و دکتر مهدی زعفرانیه		
دانشکده ریاضی و علوم کامپیوتر	رشته: علوم تصمیم و مهندسی دانش	گرایش:
مقطع: کارشناسی ارشد	تاریخ دفاع: بهمن ۱۳۹۶	تعداد صفحات: ۷۴
عنوان پایان نامه: روشی مبتنی بر گرادیان برای افزایش طیفی گراف		
کلید واژه‌ها: افزایش طیفی گراف، روش گرادیان مزدوج، روش نیستروم		
<p>چکیده: خوشه بندی داده ها از جمله مسائل مهم حوزه داده کاوی و یادگیری ماشین است که کاربردهای فراوانی در علوم مختلف دارد. یکی از روش های خوشه بندی، نمایش وابستگی داده ها توسط گراف و استفاده از شیوه های افزایش بندی گراف می باشد. خوشه بندی طیفی گراف یکی از مشهورترین این شیوه ها است که توجه زیادی را به خود معطوف نموده است. در روش خوشه بندی طیفی گراف از بردار ویژه متناظر با دومین کوچکترین مقدار ویژه برای خوشه بندی داده ها استفاده می شود. در این تحقیق دو روش تقریبی گرادیان مزدوج و نیستروم برای محاسبه مقدار ویژه و بردار ویژه بررسی شده است. شیوه ی گرادیان مزدوج، روشی تکراری است که از یک نقطه آغازین تصادفی حرکت نموده و به سمت جواب کمینه حرکت می کند و روش نیستروم، با استفاده از تعدادی نمونه از مجموعه داده، تقریبی از ماتریس مجاورت را به دست می آورد. در انتها هر دو روش پیاده سازی شده و بهبودهایی بر روی آنها انجام شده است.</p>		

پیش‌گفتار

خوشه‌بندی یکی از ابتدایی‌ترین فعالیت‌های ذهنی بشر است که برای استخراج مفهوم یا به بیانی دیگر خلاصه‌سازی اطلاعات استفاده می‌شود. خوشه‌بندی را می‌توان به عنوان مهم‌ترین روش در یادگیری بدون نظارت در نظر گرفت که در تشخیص الگوی بدون ناظر، بینایی ماشینی، تقطیع تصویر کاربرد فراوان دارد. در خوشه‌بندی هدف تقسیم کردن داده‌ها به خوشه‌ها به نحوی است که شباهت بین داده‌های هر خوشه حداکثر و تفاوت بین داده‌ها در خوشه‌های متفاوت حداقل باشد.

از روش‌های خوشه‌بندی، می‌توان خوشه‌بندی طیفی را نام برد که با استفاده از آن می‌توان مسئله پردازش تصویر و همچنین خوشه‌بندی مجموعه داده‌های بزرگ را انجام داد. از جمله کاربردهای آن می‌توان در دسته‌بندی بیمارها و تشخیص بیماری نام برد. در خوشه‌بندی طیفی، هدف افراز بهینه رئوس گراف به نحوی است که فاصله رئوس هر خوشه کمینه و فاصله بین خوشه‌های بیشینه گردد.

این پایان‌نامه شامل ۳ فصل است :

در فصل اول تعاریف و مفاهیم موردنیاز بیان خواهند شد.

در فصل دوم به بیان چند روش و الگوریتم خوشه‌بندی طیفی پرداخته خواهد شد.

در فصل سوم روش‌های نیستروم و گرادیان مزدج بر روی چند مجموعه داده اعمال و نتایج کار نمایش داده خواهد شد.

فصل ۱

تعاریف اولیه و پیش نیازها

خوشه‌بندی را می‌توان به عنوان مهم‌ترین روش در یادگیری بدون نظارت در نظر گرفت. در خوشه‌بندی هدف تقسیم کردن داده‌ها به خوشه‌ها به نحوی است که شباهت بین داده‌های هر خوشه حداکثر و تفاوت بین داده‌ها در خوشه‌های متفاوت حداقل باشد. در واقع این هدف، به‌عنوان یک معیار در انجام یک خوشه‌بندی خوب مدنظر است، یعنی نقاطی که در یک خوشه هستند مشابه با یکدیگر، و متفاوت از دیگر خوشه‌ها باشند.

فرض کنید X مجموعه‌ای از داده‌ها باشد که قصد داریم روی آنها خوشه‌بندی را انجام دهیم:

$$X = \{x_1, x_2, \dots, x_n\}$$

یک خوشه‌بندی m تایی از X ، ابزار X به m مجموعه c_1, c_2, \dots, c_m است که سه شرط زیر را دارا است:

$$c_i \neq \emptyset \quad i = 1, 2, \dots, m \quad ۱.$$

$$\bigcup_{i=1}^m c_i = X \quad ۲.$$

$$c_i \cap c_j = \emptyset, \quad i \neq j, \quad i, j = 1, 2, \dots, m \quad ۳.$$

به علاوه بردارهایی که در خوشه c_j هستند به هم دیگر شبیه‌اند و از خوشه‌های دیگر متفاوتند و هر بردار ویژگی تنها به یک خوشه تعلق دارد. بنابراین خوشه به مجموعه‌ای از داده‌ها گفته می‌شود که به هم شباهت داشته باشند.

گام‌های اساسی که یک متخصص به منظور توسعه خوشه‌بندی باید انجام دهد به شرح زیر است:

۱. انتخاب ویژگی^۱: ویژگی‌ها باید به‌طور صحیح انتخاب شوند تا توصیف درستی از الگو به‌دست دهند، یعنی اطلاعات بیشتری درخصوص آن فعالیت به‌دست آوریم و اطلاعات زائد را از میان ویژگی‌هایی که هدف اصلی ماست، حذف کنیم.

^۱Feature selection

۲. میزان نزدیکی^۱ بین الگوها: این اندازه کمیتی است که مشابه یا غیر مشابه بودن دو بردار ویژگی را نشان می دهد.
۳. محک خوشه‌بندی^۲: محک خوشه‌بندی یا تابع هزینه به تعبیری میزان معقول^۳ بودن خوشه‌ها است و از یک نگاه تابعی است که با بهینه کردن آن روی داده‌ها، داده‌ها در خوشه‌های معقول قرار می‌گیرند. اشاره به این نکته ضروری است، نمونه‌هایی با d ویژگی طبق یک محک ممکن است در خوشه‌هایی معقول قرار بگیرند درحالی که از نگاه دیگر با یک معیار یا محک دیگر آن چنان معقول نباشد. محک خوشه‌بندی می‌تواند به وسیله تابع هدف یا برخی قواعد دیگر بیان گردد.
۴. الگوریتم خوشه‌بندی: الگوریتم خوشه‌بندی به اندازه نزدیکی و محکی که انتخاب می‌کنیم، بستگی دارد. در این گام یک الگوریتم که موفق به خوشه‌بندی داده‌ها شود و مبتنی بر تابع هزینه روش می‌باشد، ارائه می‌گردد.
۵. اعتبار نتایج^۴: باید صحت و درستی نتایجی که از یک الگوریتم خوشه‌بندی به دست آمده است، بررسی شود. به عبارتی معیارهایی برای بررسی عملکرد خوشه‌بندی وجود دارد که نشان می‌دهد در روش خوشه‌بندی مورد نظر به چند جنبه خوشه‌بندی توجه شده است. مثلاً ممکن است پس از اجرای الگوریتم، یکنواختی خوشه به عنوان معیار به کار رفته اما نامتجانس بودن خوشه‌ها مورد توجه قرار نگرفته باشد.
۶. تفسیر نتایج^۵: در اکثر موارد، متخصص باید نتایج خوشه‌بندی را با دیگر شواهد تجربی به منظور به دست آوردن نتایج درست، مقایسه و تحلیل کند.

۱-۱ تاریخچه‌ای از مسأله خوشه‌بندی

خوشه‌بندی^۶ یکی از ابتدایی‌ترین فعالیت‌های ذهنی بشر است که برای استخراج مفهوم یا به بیانی دیگر خلاصه‌سازی اطلاعات استفاده می‌شود. این مفهوم در علوم مختلفی مانند علوم زیست‌شناسی و جانورشناسی، جغرافیا و زمین‌شناسی، جامعه‌شناسی و باستان‌شناسی، علوم روانپزشکی و آسیب‌شناسی، تجارت، سیستم‌های مهندسی و پردازش تصویر به کار برده شده است. خوشه‌بندی با توجه به زمینه‌ای که در آن استفاده می‌شود، نام‌های متفاوتی پیدا کرده است مانند یادگیری بدون ناظر در تشخیص الگو^۷، رده‌بندی جانوران در زیست‌شناسی، گونه‌شناسی در علوم اجتماعی و افزاز در نظریه گراف.

۱-۱-۱ کاربرد خوشه‌بندی

خوشه‌بندی در تشخیص الگوی بدون ناظر، بینایی ماشین، تقطیع تصویر^۸ کاربرد فراوان دارد. یکی از اهداف شناسایی الگو، خوشه‌بندی داده‌هاست. از جمله کاربردهای مهم تشخیص الگو در علوم پزشکی است. طی سال‌های گذشته اطلاعات

^۱Proximity measure ^۲Clustering criterion ^۳sensible ^۴Validation of the results ^۵Interpretation
of the results ^۶clustering ^۷Pattern recognition ^۸Image segmentation

بسیاری از بیماران و تشخیص نهایی و نتیجه درمان آن‌ها ذخیره شده است. الگوهای مهمی درون این اطلاعات نهفته است که از چشم انسان دور مانده است. با الگوریتم‌های خوشه‌بندی می‌توان این اطلاعات را از درون داده‌ها استخراج کرد که در تشخیص بیماری‌ها به پزشکان کمک می‌کند.

برای استفاده از خوشه‌بندی طیفی می‌توان به صورت غیر مستقیم بردار ویژه و مقدار ویژه را محاسبه کرد. دوروشی که در این تحقیق به بررسی آن‌ها پرداختیم عبارتند از:

۱. گرادیان مزدوج

۲. روش نیستروم

در روش گرادیان مزدوج به بیان یک تابع هدف برای بهینه‌سازی می‌پردازیم و در روش نیستروم با انتخاب تعدادی از داده‌ها به عنوان نمونه، از محاسبه بردارهای ویژه ماتریس‌های بزرگ صرف نظر می‌کنیم و به تقریب مناسبی از آن‌ها اکتفا می‌کنیم.

۲-۱ گراف

برای تعاریف این بخش از مرجع [۱] استفاده شده است.

تعریف ۱-۲-۱. مجموعه چندگانه^۱ مجموعه‌ای است که اعضای آن بیش از یک بار ظاهر می‌شوند، پس هر عضو در یک مجموعه چندگانه یک نرخ تکرار^۲ دارد.

تعریف ۲-۲-۱. زوج $G = (V, E)$ که در آن V مجموعه‌ای متناهی و E مجموعه‌ای از زیرمجموعه‌های دو عضوی مجموعه‌ی V است را گراف G گویند.

مجموعه‌ی V را مجموعه‌ی رئوس و اعضای آن را رئوس گراف G گویند و مجموعه‌ی E را مجموعه‌ی یال‌ها و اعضای آن را یال‌های گراف G گویند.

تعریف ۳-۲-۱. زوج $G = (V, A)$ را گراف جهت‌دار^۳ G گویند هرگاه V مجموعه‌ای متناهی و A مجموعه‌ای چندگانه از زوج‌های مرتب از اعضای مجموعه V باشد.

مجموعه‌ی V را مجموعه‌ی رئوس و اعضای آن را رئوس گراف جهت‌دار G گویند و مجموعه‌ی A را مجموعه‌ی کمان‌ها و اعضای آن را کمان‌های گراف جهت‌دار G گویند.

از نماد n برای نمایش تعداد رئوس گراف استفاده می‌کنیم. یعنی داریم: $|V| = n$

تعریف ۴-۲-۱. رأس مجاور رأس v در گراف G رأسی است که با یالی به v وصل شده باشد.

^۱multi-set ^۲multiplicity ^۳Directed graph

تعریف ۱-۲-۵. رأسی که از آن هیچ یالی عبور نکند، یک رأس منفرد (تنها) نامیده می‌شود.

تعریف ۱-۲-۶. هر یالی که از یک رأس به همان رأس متصل شود، حلقه نامیده می‌شود.

تعریف ۱-۲-۷. گراف بدون جهت، بدون حلقه و بدون یال‌های چندگانه، گراف ساده نامیده می‌شود.

تعریف ۱-۲-۸. تعداد یال‌های متصل به یک رأس، درجه آن رأس نامیده می‌شود.

تعریف ۱-۲-۹. یک گردش بین رأس‌ها که در آن، رأس تکراری مجاز نیست، مسیر نامیده می‌شود.

تعریف ۱-۲-۱۰. گراف G کامل نامیده می‌شود اگر از هر رأس به تمامی رئوس آن، یالی متصل باشد.

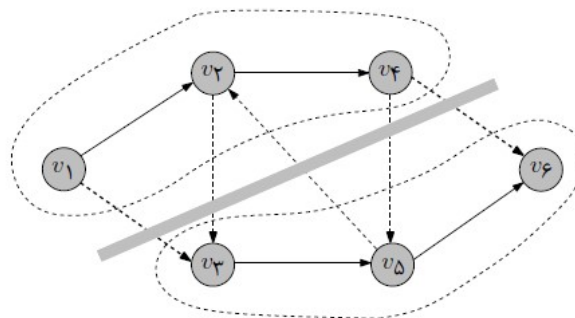
تعریف ۱-۲-۱۱. گراف H را یک زیرگراف از G گوئیم اگر مجموعه‌ی رئوس H زیرمجموعه‌ای از مجموعه رئوس G باشد و همچنین هر یال uv از H یالی از G باشد.

تعریف ۱-۲-۱۲. یک گراف را همبند^۱ گوئیم اگر به ازای هر جفت رأس u و v ، یک مسیر بدون جهت بین u و v وجود داشته باشد.

تعریف ۱-۲-۱۳. زیرگراف H از G را یک مؤلفه از G گوئیم، هرگاه H همبند باشد و زیرگراف همبندی از G به غیر از H وجود نداشته باشد که H زیرگراف آن باشد.

تعریف ۱-۲-۱۴. افزاز مجموعه رئوس یک گراف به دو زیرمجموعه ناتهی S و $\bar{S} = V - S$ را یک برش گوئیم و آن را با $[S, \bar{S}]$ نمایش می‌دهیم.

تعریف ۱-۲-۱۵. برای دو رأس s و t از گراف، برش $[S, \bar{S}]$ را یک st -برش گوئیم هرگاه $s \in S$ و $t \in \bar{S}$ باشد. در شکل ۱-۱ مفهوم برش نشان داده شده است.



شکل ۱-۱: یک v_1v_6 برش با $S = \{v_1, v_2, v_4\}$ و $\bar{S} = \{v_3, v_5, v_6\}$

تعریف ۱-۲-۱۶. یک برش از یک گراف همبند را یک برش ساده گوئیم اگر گراف حاصل از حذف کمان‌های متناظر با آن برش دقیقاً دو مؤلفه داشته باشد.

^۱connected

تعریف ۱-۲-۱۷. ماتریس مربعی که تمام درایه‌های خارج از قطر اصلی آن صفر باشد را ماتریس قطری می‌گوییم.

تعریف ۱-۲-۱۸. فرض کنید $V(G) = \{v_1, v_2, \dots, v_n\}$ ، ماتریس مجاورت^۱ گراف G را با $W_{n \times n}$ نمایش می‌دهیم که درایه i, j ام آن برابر است با تعداد یال‌هایی که v_i را به v_j متصل می‌کند. بدیهی است در گراف‌های بدون جهت به ازای هر i و j

$$W_{ij} = \begin{cases} 1 & \text{اگر } v_i \text{ به } v_j \text{ وصل باشد} \\ 0 & \text{در غیراین صورت} \end{cases}$$

و $W_{ij} = W_{ji}$ ، یعنی ماتریس مجاورت یک گراف بدون جهت، متقارن است.

اگر گراف بدون حلقه باشد، به عبارت دیگر اگر داشته باشیم $W_{ii} = 0, \forall i$ ، در این صورت درایه‌های روی قطر اصلی ماتریس مجاورت، همگی صفر می‌باشند.

لذا به راحتی مشاهده می‌شود، ماتریس مجاورت یک گراف ساده، یک ماتریس متقارن با درایه‌های صفر و یک می‌باشد که تمام درایه‌های قطر اصلی آن الزاماً صفر است.

تعریف ۱-۲-۱۹. فرض کنید $V(G) = \{v_1, \dots, v_n\}$. ماتریس درجه D گراف بدون جهت G ، ماتریسی قطری است که عناصر روی قطر اصلی d_{ii} ، درجه رأس i ام است.

۳-۱ ماتریس

تعاریف و قضایای این بخش تا تعریف ۱-۳-۱۳ از مرجع [۲] اقتباس شده‌اند.

تعریف ۱-۳-۱. ترانهاد ماتریس A با A^T نشان داده می‌شود که در آن $[A^T]_{ji} = [A]_{ij}$.

تعریف ۱-۳-۲. ماتریس A را قطری شدنی می‌گوییم اگر ماتریس وارون پذیر P موجود باشد به طوری که $P^{-1}AP$ یک ماتریس قطری باشد.

تعریف ۱-۳-۳. ماتریس $A \in M_n$ نرمال نامیده می‌شود اگر $AA^T = A^T A$.

تعریف ۱-۳-۴. ماتریس $A \in \mathbb{C}^{n \times n}$ هرمیتی نامیده می‌شود اگر ترانهاد مزدوج مختلط آن با خودش برابر باشد و $a_{ij} = a_{ji}^*$ که با نماد $A = A^\dagger$ نشان داده می‌شود (* نشان‌دهنده مزدوج عدد مختلط است).

تعریف ۱-۳-۵. ماتریس $B \in M_n$ متقارن نامیده می‌شود اگر $b_{ij} = b_{ji}$ که با نماد $B = B^T$ نشان داده می‌شود. ماتریس متقارن صورت خاصی از ماتریس هرمیتی است که تمام درایه‌هایش اعداد حقیقی باشند.

^۱adjacency matrix

تعریف ۱-۳-۶. رتبه^۱ ماتریس A ، بعد فضای برداری تولید شده (اسپن شده) با ستون‌های آن است که متناظر با بیشترین تعداد ستون‌های مستقل خطی A است. که با بعد فضای تولید شده توسط سطرهایش برابر است.

تعریف ۱-۳-۷. ماتریس $A \in M_n$ معین مثبت نامیده می‌شود هرگاه به ازای هر بردار $x \in \mathbb{R}^n$ داشته باشیم:

$$x^T A x > 0$$

تعریف ۱-۳-۸. ماتریس $A \in M_n$ نیمه معین مثبت نامیده می‌شود هرگاه به ازای هر بردار $x \in \mathbb{R}^n$ داشته باشیم:

$$x^T A x \geq 0$$

تعریف ۱-۳-۹. افزایش سطرها و ستونهای ماتریس $A_{m \times n}$ بلوکی کردن ماتریس A نامیده می‌شود و هر یک از افزایش ایجاد شده یک بلاک یا زیر ماتریس برای ماتریس A نامیده می‌شود.

تعریف ۱-۳-۱۰. اگر V یک فضای برداری روی میدان F باشد مجموعه $S = \{x_1, \dots, x_m\} \subseteq V$ را مستقل خطی می‌نامیم اگر

$$\lambda_i \in \mathbb{R} : \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_m x_m = 0 \Rightarrow \lambda_1 = \lambda_2 = \dots = \lambda_m = 0$$

و S را وابسته خطی می‌نامیم اگر

$$\exists (\lambda_1, \dots, \lambda_m) \neq (0, \dots, 0) \Rightarrow \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_m x_m = 0$$

تعریف ۱-۳-۱۱. ماتریس دلخواه $M_{n \times n}$ را در نظر بگیرید. بردار غیر صفر v یک بردار ویژه با مقدار ویژه λ برای ماتریس M خواهد بود اگر رابطه $Mv = \lambda v$ برقرار باشد. هر ماتریس $M_{n \times n}$ دارای n بردار ویژه و n مقدار ویژه است.

تعریف ۱-۳-۱۲. تعداد دفعات تکرار یک مقدار ویژه به عنوان ریشه معادله مشخصه ماتریس را تکرر جبری^۲ آن مقدار ویژه می‌نامیم.

تعریف ۱-۳-۱۳. برای ماتریس $M_{n \times n}$ با مقادیر ویژه $\lambda_1, \lambda_2, \dots, \lambda_n$ اثر ماتریس M به صورت زیر محاسبه می‌شود:

$$\text{trace}(M) = \lambda_1 + \lambda_2 + \dots + \lambda_n$$

^۱rank ^۲Algebraic multiplicity

تعریف ۱-۳-۱۴. حاصلضرب داخلی دو بردار $x = (x_1, \dots, x_n)^T$ و $y = (y_1, \dots, y_n)^T$ بصورت زیر تعریف می شود [۳]:

$$x \cdot y = x^T y = \sum_{i=1}^n x_i y_i$$

حاصلضرب داخلی دو بردار x و y را با $\langle x, y \rangle$ نیز نمایش می دهند.

تعریف ۱-۳-۱۵. بردارهای متعامد: $v_1, \dots, v_m \in \mathbb{R}^n$

$$(v_i \cdot v_j) = 0 \quad \forall i \neq j, \quad \|v_i\| = \sqrt{v_i \cdot v_i} = 1$$

تعریف ۱-۳-۱۶. ماتریس $V \in \mathbb{R}^{n \times n}$ را متعامد گویند هرگاه:

$$V V^T = I$$

حداکثر n بردار متعامد در \mathbb{R}^n می توان یافت.

اگر یک ماتریس متعامد باشد آنگاه سطرهایش (ستونهایش) هم متعامدند.

تعریف ۱-۳-۱۷. پایه: n بردار متعامد، یک پایه برای \mathbb{R}^n تشکیل می دهند.

تعریف ۱-۳-۱۸. اگر X ماتریس شامل بردارهای ستونی $\{x_1, \dots, x_m\}$ باشد، ماتریس $G = X^T X$ ماتریس گرام نامیده می شود و درایه های آن برابر است با حاصلضرب داخلی بردارهای $\{x_1, \dots, x_m\}$ یعنی $G_{ij} = x_i^T x_j$. ماتریس گرام متقارن و نیمه معین مثبت است.

تعریف ۱-۳-۱۹. فرض کنید A و B دو ماتریس $n \times n$ باشند. λ را یک مقدار ویژه تعمیم یافته زوج (A, B) می نامیم هرگاه بردار $x \neq 0$ یافت شود به طوری که $Ax = \lambda Bx$. بردار x بردار ویژه تعمیم یافته زوج (A, B) نامیده می شود و معادله مشخصه آن بصورت زیر است [۴]:

$$(A - \lambda B)x = 0$$

لم ۱-۳-۲۰. اگر $M_{n \times n}$ یک ماتریس متقارن باشد و x و y دو بردار ویژه متناظر با مقادیر ویژه λ و μ ماتریس M که $\lambda \neq \mu$ آنگاه، x و y متعامدند.

برهان. برای هر ماتریس حقیقی A و هر بردار x و y رابطه زیر برقرار است:

$$\langle Ax, y \rangle = \langle x, A^T y \rangle$$

اکنون فرض می‌کنیم M یک ماتریس متقارن است و x و y بردارهای ویژه ماتریس M به ترتیب متناظر با مقادیر ویژه متمایز λ و u می‌باشند، با توجه به رابطه فوق داریم:

$$\lambda \langle x, y \rangle = \langle \lambda x, y \rangle = \langle Mx, y \rangle = \langle x, M^T y \rangle = \langle x, My \rangle = \langle x, uy \rangle = u \langle x, y \rangle$$

بنابراین

$$(\lambda - u) \langle x, y \rangle = 0$$

و چون $\lambda - u \neq 0$ پس داریم:

$$\langle x, y \rangle = 0 \quad (x \perp y)$$

□ پس نتیجه می‌گیریم که بردارهای ویژه یک ماتریس متقارن متعامدند.

تعریف ۱-۳-۲۱. پایه متعامد یک^۱: سطرهای ماتریس لم ۱-۳-۲۰ یک پایه متعامد یکه برای فضای \mathbb{R}^n اند.

قضیه ۱-۳-۲۲. (قضیه طینی برای ماتریس‌های هرمیتی) فرض کنید M یک ماتریس $n \times n$ هرمیتی با مقادیر حقیقی یا مختلط، آنگاه مقادیر ویژه $\lambda_1, \dots, \lambda_n$ آن حقیقی‌اند و یک پایه متعامد یکه از بردارهای ویژه v_1, \dots, v_n دارد. علاوه بر این $M = \sum_{i=1}^n \lambda_i v_i v_i^\dagger$.

یک اثبات جزئی برای قضیه آورده می‌شود. نشان داده می‌شود که تمام مقادیر ویژه حقیقی‌اند. این بردارهای ویژه متناظر با مقادیر ویژه مجزا متعامدند. اگر تمام مقادیر ویژه مجزا باشند که اثبات کامل است. در غیر این صورت، اثباتی در ادامه برای آن می‌آوریم. اگر ماتریس حقیقی باشد، می‌توانیم تمام بردارهای ویژه را حقیقی بدانیم و مزدوج ماتریس با ترانهاد هم ارز است، اگرچه ما فقط با ماتریس‌های حقیقی کار می‌کنیم، اما اثبات کلی‌تری را ارائه می‌دهیم، زیرا دقیقاً یک ساختار مشابه دارند:

برهان. فرض می‌کنیم که λ یک مقدار ویژه M و v بردار ویژه متناظر با آن باشد، بنابراین $Mv = \lambda v^T$. پس $v^T(Mv) = \lambda v^T v$ از طرفی

$$v^\dagger(Mv) = (v^\dagger M)v = (v^\dagger M^\dagger)v = (Mv)^\dagger v = (\lambda v)^\dagger v = \lambda^* v^\dagger v$$

، چون $M = M^\dagger$ ماتریس هرمیتی است $M = M^\dagger$. نشان دادیم که $\lambda v = \lambda^* v$ بنابراین $\lambda = \lambda^*$ که نتیجه می‌گیریم λ حقیقی است. (بردار صفر به عنوان یک بردار ویژه در نظر گرفته نشده است).

^۱orthonormal basis (ONB)

فرض می‌کنیم v_1 و v_2 بردارهای ویژه متناظر با مقادیر ویژه مجزا λ_1 و λ_2 اند. بنابراین $v_2^\dagger M v_2 = \lambda_2 v_2^\dagger v_2 = \lambda_2$ و $v_1^\dagger (M v_2) = v_1^\dagger \lambda_2 v_2 = \lambda_2 v_1^\dagger v_2 = 0$. از طرف دیگر با استفاده از $M = M^\dagger$ داریم:

$$(v_1^\dagger M) v_2 = (v_1^\dagger M^\dagger) v_2 = ((M v_1)^\dagger) v_2 = (\lambda_1 v_1)^\dagger v_2 = \lambda_1^* v_1^\dagger v_2.$$

پس، $\lambda_2 v_1^\dagger v_2 = \lambda_1^* v_1^\dagger v_2$ ، که نتیجه می‌دهد $(\lambda_2 - \lambda_1^*) v_1^\dagger v_2 = 0$. اما از آنجایی که تمام بردارهای ویژه حقیقی اند پس $\lambda_1^* = \lambda_1$ و چون طبق فرض λ_1 از λ_2 مجزا است می‌رسیم به $v_1^\dagger v_2 = 0$ به این معنی که v_1 و v_2 با یکدیگر متعامدند. سرانجام، فرض می‌کنیم که تمام n مقادیر ویژه $\lambda_1, \dots, \lambda_n$ مجزا هستند، پس بردارهای ویژه v_1, \dots, v_n متناظرشان متقابلاً متعامدند که نتیجه می‌دهد مستقل خطی اند. پس یک پایه برای \mathbb{R}^n تشکیل می‌دهند. بدون از دست دادن کلیت مسئله، ممکن است فرض کنیم که تمامشان (بردارهای ویژه) نرم واحد دارند که در نتیجه یک پایه متعامد یک‌به‌یک بدست می‌آوریم.

حال قسمت دوم قضیه را اثبات می‌کنیم. نشان می‌دهیم که برای تمام $v \in \mathbb{R}^n$ تساوی $Mv = \sum_{i=1}^n \lambda_i v_i v_i^\dagger v$ برقرار است. به وضوح، کفایت می‌کند که نشان دهیم این تساوی برای تمام v ها در پایه‌هایی از \mathbb{R}^n حفظ می‌شود. با در نظر گرفتن مطالبی که در بالا برای بردارهای ویژه گفتیم، داریم $Mv_j = \lambda_j v_j$ ، نظر به این که

$$\sum_{i=1}^n \lambda_i v_i v_i^\dagger v_j = \sum_{i=1}^n \lambda_i v_i (v_i^\dagger v_j) = \lambda_j v_j (v_j^\dagger v_j) = \lambda_j v_j.$$

برای بدست آوردن تساوی دوم بالا از این حقیقت استفاده کردیم که، از آنجایی که بردارهای v_i ، $i = 1, \dots, n$ متقابلاً متعامدند، $v_i^\dagger v_j = 0$ هر جا که $i \neq j$. آخرین تساوی از $\|v_j\|^2 = 1$ بدست می‌آید. \square

قضیه ۱-۳-۲۳. فرض کنیم $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ نشان‌دهنده مقادیر ویژه ماتریس $M \in \mathbb{R}^{n \times n}$ باشند. آنگاه:

$$\lambda_1 = \min_{x \in \mathbb{R}^n \setminus \{0\}} \frac{x^T M x}{x^T x}, \quad \lambda_n = \max_{x \in \mathbb{R}^n \setminus \{0\}} \frac{x^T M x}{x^T x},$$

که λ_1 نشان‌دهنده بردار صفر در \mathbb{R}^n است. علاوه بر این، اگر v_1, \dots, v_n یک پایه متعامد یک‌به‌یک از بردارهای ویژه M هم ارز با مقادیر ویژه $\lambda_1 \dots \lambda_n$ باشند، آنگاه

$$\lambda_{k+1} = \min_{x \perp v_1, \dots, v_k} \frac{x^T M x}{x^T x},$$

به عنوان مثال، مینیمم متجاوز از بردارهای غیر صفر گرفته شده است که با تمام k بردارهای ویژه اول متعامد است (و به همین ترتیب با k زیرفضای تولید شده توسط آنها).

تقسیم $x^T M x / x^T x$ به عنوان خارج قسمت رایلی^۱ یا تقسیم رایلی-ریتر^۲ شناخته می‌شود. مشابه قسمت آخر قضیه،

^۱Rayleigh quotient

^۲Rayleigh-Ritz

λ_k می‌تواند به عنوان ماکسیمم خارج قسمت رایلی روی فضای تولید شده بردارهای ویژه متناظر با $n - k$ بزرگترین مقادیر ویژه بدست آید.

برهان. از آنجایی که بردارهای ویژه M ، v_1, \dots, v_n یک پایه متعامد یکه تشکیل می‌دهند، هر $x \in \mathbb{R}^n$ می‌تواند از حاصلضرب خطی آن‌ها بدست آید، $x = \sum_{i=1}^n a_i v_i$. علاوه بر این $a_i = x^T v_i$ در نتیجه با استفاده از تجزیه طیفی M از قضیه ۱-۳-۲۲ داریم:

$$x^T M x = \sum_{i=1}^n x^T \lambda_i v_i v_i^T x = \sum_{i=1}^n \lambda_i a_i^2,$$

از آنجایی که:

$$x^T x = \sum_{i,j=1}^n a_i a_j v_i^T v_j = \sum_{i=1}^n a_i^2.$$

اما $\sum_{i=1}^n \lambda_i a_i^2 \geq \sum_{i=1}^n \lambda_1 a_i^2$ ، پس با ادامه تساوی بالا برای هر $x \in \mathbb{R}^n$ می‌رسیم به $x^T M x \geq \lambda_1 x^T x$. علاوه بر این، تساوی بر $x = v_1$ حفظ می‌شود.

حکم دوم قضیه به طور مشابه اثبات می‌شود. با مشاهده اینکه $\sum_{i=1}^n \lambda_i a_i^2 \leq \sum_{i=1}^n \lambda_n a_i^2$ ، و از آنجایی که برای تمام $x \in \mathbb{R}^n$ داریم $x^T M x \geq \lambda_1 x^T x$ با تساوی برای $x = v_n$.

اثبات قسمت آخر دقیقاً مشابه است، با دانستن اینکه اگر $x \in \mathbb{R}^n$ با بردارهای v_1, \dots, v_n متعامد باشند، آنگاه ضرایب a_1, \dots, a_k تمامشان برابر با صفر می‌شوند. از آنجایی که

$$\sum_{i=1}^n \lambda_i a_i^2 = \sum_{i=k+1}^n \lambda_i a_i^2 \geq \sum_{i=k+1}^n \lambda_{k+1} a_i^2.$$

□

باقی اثبات یکسان است.

۴-۱ نُرم

تعریف ۱-۴-۱. تابع حقیقی $\|\cdot\|$ تعریف شده بر روی فضای برداری V را نُرم نامیم اگر در سه خاصیت زیر صدق کند [۵]:

۱. به ازای هر $x \in V$ ، $\|x\| \geq 0$ و $\|x\| = 0$ اگر و فقط اگر $x = 0$.

۲. به ازای هر $x \in V$ و $\alpha \in \mathbb{R}$ ، $\|\alpha x\| = |\alpha| \|x\|$.

۳. به ازای هر $x, y \in V$ (نابرابری مثلثی) $\|x + y\| \leq \|x\| + \|y\|$.

فضای برداری V مجهز به نرم $\|\cdot\|$ را یک فضای برداری نرم‌دار می‌نامیم. از آنجایی که دامنه تعریف نرم، فضای برداری است، بسته به اینکه اعضای فضای برداری چه باشند، نرم ممکن است برای بردار، ماتریس، یا تابع، تعریف شود. ورودی نرم، عضوهای فضای برداری و خروجی آن عدد حقیقی مثبتی است پس هر نرم، مجموعه اعداد حقیقی مثبت می‌باشد. می‌توانیم در مورد $\|x\|$ به عنوان طول یا اندازه بردار x فکر کنیم. یک نرم بر روی یک فضای برداری مفهوم قدر مطلق، $|r|$ ، برای یک عدد حقیقی یا اندازه برای عدد مختلط را تعمیم می‌دهد.

۱-۴-۱ انواع نرم

معروف‌ترین نرم‌ها در \mathbb{R}^n نرم اقلیدسی l_2 است که به صورت زیر تعریف می‌شود [۵]:

$$\|x\|_2 = \left(\sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}}$$

که در آن $x = (x_1, x_2, \dots, x_n)^T$. این نرمی است که متناظر با مفهوم ذاتی طول می‌باشد. نوع دیگر نرم، نرم بی‌نهایت l_∞ است که به صورت زیر تعریف می‌شود:

$$\|x\|_\infty = \max |x_i|$$

سومین نرم مهم در \mathbb{R}^n ، نرم l_1 نامیده می‌شود:

$$\|x\|_1 = \sum_{i=1}^n |x_i|$$

در حالت کلی داریم:

$$\|x\|_p = \left(\sum_{i=1}^n x_i^p \right)^{\frac{1}{p}}$$

۵-۱ نرم‌های ماتریسی

حال به تعریف نرم برای ماتریس‌ها می‌پردازیم. اگر چه ما با نرم‌های ماتریسی سروکار داریم که شرایط آنها، همان شرایط (۱) تا (۳) می‌باشند، اما معمولاً نرم ماتریسی را ترجیح می‌دهیم که رابطه معنی داری با یک نرم برداری داشته باشد و اگر یک نرم

بردارى $\|\cdot\|$ مشخص شده باشد نرم ماتریسی تبعی (طبیعی) به صورت

$$\|A\| = \sup\{\|Au\| : u \in \mathbb{R}^n, \|u\| = 1\} \quad (1-1)$$

تعریف می‌گردد. این نرم ماتریسی وابسته به نرم برداری مفروض نیز نامیده می‌شود. در اینجا A یک ماتریس $n \times n$ است [۶].

قضیه ۱-۵-۱. اگر $\|\cdot\|$ نرمی در \mathbb{R}^n باشد، آنگاه رابطه (۱-۱) یک نرم بر روی فضای خطی همه ماتریس‌های $n \times n$ تعریف می‌کند.

برای اثبات به مرجع [۶] مراجعه کنید.

اگر نرم بردار $\|\cdot\|_\infty$ به صورت

$$\|x\|_\infty = \max |x_i|$$

تعریف شود، آنگاه نرم ماتریسی طبیعی آن به صورت زیر تعریف می‌شود:

$$\|A\|_\infty = \max \sum_{j=1}^n |a_{ij}| \quad , \quad 1 \leq i \leq n$$

نرم ماتریسی مهم دیگر، نرم ماتریسی l_2 به نام نرم طیفی است که به صورت زیر تعریف می‌شود:

$$\|A\|_2 = \sup \|Ax\|_2 = \max |\sigma_i|$$

که در آن σ_i ها مقادیر تکین A می‌باشند. در اینجا نرم برداری طبیعی، نرم اقلیدسی است.

نرم فربنیوس ماتریسی هم به صورت زیر تعریف می‌شود:

$$\|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2}$$

۶-۱ تجزیه مقدار تکین

تعریف ۱-۶-۱. ماتریس $A \in \mathbb{R}^{m \times n}$ را شبه قطری گویند هر گاه

$$a_{i,j} = 0 \quad \text{اگر} \quad i \neq j$$

تعریف ۱-۶-۲. فرض کنید A یک ماتریس حقیقی $m \times n$ باشد. تعریف می‌شود

$$B = A^T A$$

ماتریس B مربعی مرتبه n و نیمه معین مثبت است، زیرا برای هر $x \in \mathbb{R}$ و $x \neq 0$ داریم

$$x^T B x = x^T A^T A x = (Ax)^T (Ax) = \|Ax\|^2 \geq 0$$

پس مقادیر ویژه B نامنفی‌اند. فرض کنید مقادیر ویژه B ، $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ باشند، حال تعریف می‌شود $\sigma_i = \sqrt{\lambda_i}$ و $i = 1, 2, \dots, n$ ها مقادیر تکین A نامیده می‌شوند. واضح است که

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$$

در تجزیه مقدار تکین A یک ماتریس $A \in \mathbb{R}^{m \times n}$ ، به حاصلضرب سه ماتریس U ، Σ ، V تبدیل می‌شود به طوری

که

$$A = U \Sigma V^T = [u_1, u_2, \dots, u_m] \begin{bmatrix} D & \vdots & O_1 \\ \dots & \dots & \dots \\ O_2 & \vdots & O_3 \end{bmatrix} [v_1, v_2, \dots, v_n]^T$$

که در آن U و V ماتریس‌های یکا متعامد و به ترتیب با ابعاد $m \times m$ و $n \times n$ می‌باشند و ماتریس Σ یک ماتریس شبه قطری $m \times n$ است. O_1 ، O_2 و O_3 ماتریس‌های صفر می‌باشند و D یک ماتریس قطری است که عناصر روی قطر آن مقادیر تکین ماتریس A هستند. علاوه بر این عناصر روی قطر D نامنفی هستند و داریم $D_{ii} = \sigma_i$ ، زیرا σ_i ها جذر مقادیر ویژه ماتریس $A^T A$ یا $A A^T$ هستند و نمی‌توانند مقدار منفی بگیرند. همچنین این مقادیر به ترتیب نزولی روی قطر اصلی ماتریس D مرتب شده‌اند و داریم [۵]:

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq \sigma_p \geq 0$$

که $p = \min(m, n)$

^۱Singular value

^۲Singular-value decomposition

۷-۱ تقریب مرتبه پایین

تقریب مرتبه پایین^۱ ماتریس داده شده $A \in \mathbb{R}^{m \times n}$ برابر با ماتریس $\hat{A} \in \mathbb{R}^{m \times n}$ است به گونه‌ای که $\text{rank}(\hat{A}) = r$ و $r \leq \min(m, n)$ باشد. استفاده از تجزیه مقدار تکین بهترین تقریب را برای این مسئله نتیجه می‌دهد [۷]. تعریف ریاضی مسئله به صورت زیر است:

$$\text{minimize over } \hat{A} \quad \|A - \hat{A}\|_F \quad \text{subject to} \quad \text{rank}(\hat{A}) \leq r \quad (۲-۱)$$

فرض کنید

$$A = U \Sigma V^T \in \mathbb{R}^{m \times n}, \quad m \leq n \quad (۳-۱)$$

تجزیه مقدار تکین ماتریس A باشد و U ، $\Sigma =: \text{diag}(\sigma_1, \dots, \sigma_m)$ و V به صورت زیر افراز بندی شده باشند:

$$U =: \begin{bmatrix} U_1 & U_r \end{bmatrix}, \quad \Sigma =: \begin{bmatrix} \Sigma_1 & \circ \\ \circ & \Sigma_r \end{bmatrix}, \quad V =: \begin{bmatrix} V_1 & V_r \end{bmatrix}$$

جایی که Σ_1 که $r \times r$ باشد، U_1 ، $m \times r$ و V_1 و $n \times r$ باشند. آنگاه ماتریس رتبه- r از ضرب زیر به دست می‌آید:

$$\hat{A} = U_1 \Sigma_1 V_1^T$$

به گونه‌ای که

$$\|A - \hat{A}\|_F = \min_{\text{rank}(\hat{A}) \leq r} \|A - \hat{A}\|_F = \sqrt{\sigma_{r+1}^2 + \dots + \sigma_m^2} \quad (۴-۱)$$

\hat{A} یکتاست اگر و تنها اگر $\sigma_r \neq \sigma_{r+1}$.

مثال ۷-۱-۱. ماتریس $A \in \mathbb{R}^{4 \times 3}$ زیر را در نظر بگیرید.

$$A = \begin{bmatrix} 9 & 6 & 6 \\ 3 & 5 & 5 \\ 2 & 4 & 9 \\ 3 & 8 & 3 \end{bmatrix}$$

^۱Low-rank approximation

رتبه این ماتریس برابر با ۳ است. تجزیه مقدار تکین این ماتریس به صورت زیر است:

$$A = U\Sigma V^T$$

$$U = \begin{bmatrix} -0,6355 & 0,4450 & 0,6183 & 0,1257 \\ -0,4097 & -0,0999 & -0,1681 & -0,8910 \\ -0,4835 & -0,8187 & 0,0297 & 0,3084 \\ -0,4411 & 0,3489 & -0,7672 & 0,3084 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 18,6189 & 0 & 0 \\ 0 & 5,4569 & 0 \\ 0 & 0 & 4,3079 \\ 0 & 0 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} -0,4962 & 0,5708 & 0,6542 \\ -0,6082 & 0,3092 & -0,7311 \\ -0,6196 & -0,7607 & 0,1938 \end{bmatrix}$$

حاصل ضرب $U\Sigma V^T$ برابر با ماتریس A می شود که رتبه ۴ دارد و به وضوح

$$\|A - A\|_F = 0$$

با صفر قرار دادن سومین مقدار تکین در ماتریس Σ و استفاده از دو مقدار تکین، ماتریس \hat{A} زیر حاصل می شود که رتبه ۲ دارد.

$$\hat{A} = U \begin{bmatrix} 18,6189 & 0 & 0 \\ 0 & 5,4569 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} V^T = \begin{bmatrix} 7,2575 & 7,9473 & 5,4839 \\ 3,4737 & 4,4706 & 5,1403 \\ 1,9164 & 4,0934 & 8,9752 \\ 5,1621 & 5,5838 & 3,6403 \end{bmatrix}$$

و

$$\|A - \hat{A}\|_F = 4,3079$$

که به ازای تمام ماتریس های رتبه ۲ این مقدار کمینه است. برای به دست آوردن مقدار کمینه برای ماتریس های رتبه ۱ کفایت تنها از مقدار تکین اول در Σ استفاده کنیم.

۸-۱ ماتریس لاپلاسین گراف

اصلی ترین ابزار خوشه بندی طیفی، ماتریس لاپلاسین گراف است [۸]. الگوریتم های خوشه بندی طیفی عموماً بر پایه استفاده از این ماتریس ها و تجزیه ویژه و استفاده از بردارها و مقادیر ویژه آنها برای تحلیل دانش موجود در ساختار داده ها می باشند. تئوری طیفی گراف [۹] عنوانی است که به این موضوع می پردازد.

تعریف ۱-۸-۱. فرض کنید $G = (V, E)$ یک گراف ساده باشد که n رأس دارد و W ماتریس مجاورت و D ماتریس درجه گراف باشد، ماتریس لاپلاسین $L(G)$ متناظر با گراف G یک ماتریس $n \times n$ متقارن است و به صورت زیر تعریف می شود:

$$L = D - W \quad (۵-۱)$$

و مقادیر درایه های آن بصورت زیر می باشند:

$$L_{ij}(G) = \begin{cases} d_i & \text{if } i = j \\ -1 & \text{if } i \neq j \quad (v_i, v_j) \in E \quad \forall i, j = 1, \dots, n \\ 0 & \text{if } i \neq j \quad (v_i, v_j) \notin E \quad \forall i, j = 1, \dots, n \end{cases}$$

که d_i درجه راس i ام گراف باشد.

قضیه ۱-۸-۲. ویژگیهای ماتریس لاپلاسین عبارتند از [۸۰]:

۱. برای هر بردار $f = [f_1, f_2, \dots, f_n]^T \in \mathbb{R}^n$ داریم:

$$f^T L f = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n W_{ij} (f_i - f_j)^2$$

۲. L ماتریس متقارن و نیمه معین مثبت است.

۳. کوچکترین مقدار ویژه ماتریس L برابر است با صفر و متناظر است با بردار ویژه \vec{j} .

برهان. ۱. با توجه به تعریف L ، داریم:

$$\begin{aligned}
f^T L f &= f^T D f - f^T W f \\
&= \sum_{i=1}^n d_i f_i^2 - \sum_{i=1}^n \sum_{j=1}^n f_i f_j W_{ij} \\
&= \frac{1}{2} \left(\sum_{i=1}^n d_i f_i^2 - 2 \sum_{i=1}^n \sum_{j=1}^n f_i f_j W_{ij} + \sum_{j=1}^n d_j f_j^2 \right) \\
&= \frac{1}{2} \left(\sum_{i=1}^n \sum_{j=1}^n W_{i,j} f_i^2 - 2 \sum_{i=1}^n \sum_{j=1}^n f_i f_j W_{ij} + \sum_{i=1}^n \sum_{j=1}^n W_{i,j} f_j^2 \right) \\
&= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n W_{i,j} (f_i - f_j)^2
\end{aligned}$$

۲. متقارن بودن ماتریس L ، مستقیماً از متقارن بودن ماتریس‌های D و W نتیجه می‌شود. برای نیمه‌معین مثبت بودن یک ماتریس باید شرط $f^T L f \geq 0$ برای هر بردار دلخواه $f \in \mathbb{R}^n$ برقرار باشد. با توجه به اثبات قسمت (۱) و نامنفی بودن W_{ij} ها، به‌وضوح این شرط برای ماتریس L برقرار است.

۳. با توجه به نیمه‌معین مثبت بودن ماتریس L واضح است که $\forall f \in \mathbb{R}^n : f^T L f \geq 0$. اگر فرض کنیم x یک بردار ویژه با مقدار ویژه λ برای ماتریس L باشد، داریم:

$$Lx = \lambda x \implies x^T Lx = \lambda x^T x$$

بنابراین با توجه به اینکه $x^T Lx \geq 0$ و $x^T x > 0$ نتیجه می‌گیریم که $\lambda \geq 0$ و $\lambda = 0$ یک مقدار ویژه متناظر با بردار ویژه $\vec{j} = [1, 1, \dots, 1]^T$ است زیرا

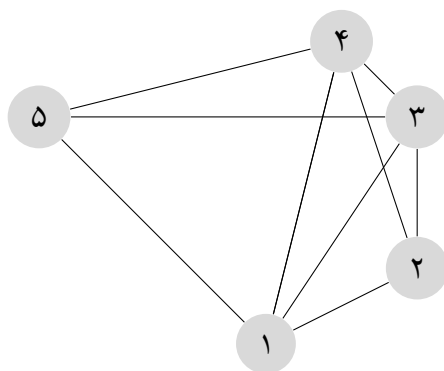
$$L\vec{j} = (D - W)\vec{j} = D\vec{j} - W\vec{j} = D - D = 0$$

بنابراین کوچکترین مقدار ویژه ماتریس لاپلاسیان، همیشه صفر است و بردار ویژه متناظر با آن، بردار \vec{j} می‌باشد. \square

یکی از ویژگیهای مهمی که ماتریس لاپلاسیان و بردارها و مقادیر ویژه آن در خوشه‌بندی طیفی دارد، به‌دست آوردن تعداد مولفه‌های همبند گراف با استفاده از طیف L می‌باشد. قضیه زیر این ویژگی را مطرح می‌کند.

قضیه ۱-۸-۳. فرض کنید G یک گراف وزن دار بدون جهت با اوزان نامنفی باشد، آنگاه تعداد اجزاء همبند گراف G برابر با تکرر جبری مقدار ویژه صفر ماتریس لاپلاسیان گراف G است.

برهان. تعداد اجزاء همبند گراف G را با k نمایش می‌دهیم. ابتدا فرض می‌کنیم $k = 1$ است در این حالت گراف همبند است. همچنین فرض کنید f یک بردار ویژه متناظر با مقدار ویژه صفر ماتریس لاپلاسیان باشد. با توجه به بخش اول قضیه



شکل ۱-۲: گراف با ۵ رأس

۱-۸-۲ و رابطه ۱-۵ می توان نوشت:

$$o = f^T L f = \sum_{i=1}^n \sum_{j=1}^n W_{i,j} (f_i - f_j)^2$$

با توجه به اینکه وزن های $W_{i,j}$ نامنفی هستند، مجموع فوق زمانی می تواند صفر باشد که تمام جمله های $W_{i,j}(f_i - f_j)^2$ صفر باشند. بنابراین، اگر دو رأس v_i و v_j متصل باشند ($W_{i,j} > 0$)، آنگاه باید f_i و f_j برابر باشند. به عبارت دیگر f باید برای تمام رئوس که از طریق مسیری به هم متصلند، ثابت باشد. بنابراین در گرافی که تنها از یک جزء متصل تشکیل شده است، فقط بردار ویژه $\vec{1}$ متناظر با مقدار ویژه صفر را خواهیم داشت که بردار شاخص این جزء متصل خواهد بود. حال فرض کنید گراف از $n \geq k$ جزء همبند تشکیل شده باشد. بدون از دست دادن کلیت مسئله فرض کنید که رئوس گراف به ترتیب درون اجزاء همبند مرتب شده قرار دارند. در این حالت ماتریس W و L بلاک قطری می باشند.

$$L = \begin{pmatrix} L_1 & & O \\ & \ddots & \\ O & & L_k \end{pmatrix}$$

هر یک از بلاک های L_i خود یک ماتریس لاپلاسیان گراف است که مرتبط با زیرگراف i امین جزء همبند می باشد. برای ماتریس های بلاک-قطری مانند L در اینجا، طیف گراف از اجتماع بردارهای ویژه بلاک های L_i که در قسمتی که مرتبط با بلاک دیگری است صفر قرار داده شده باشد، به دست می آید. با توجه به اینکه هر L_i خود یک لاپلاسیان مرتبط با یک جزء همبند در گراف است و با توجه به بخش اول اثبات، هر L_i یک مقدار ویژه صفر را خواهد داشت و بردار ویژه متناظر با آن بردار $\vec{1}$ خواهد بود.

بنابراین تعداد مقدار ویژه صفر ماتریس لاپلاسیان $L(G)$ برابر تعداد مولفه های همبندی گراف G است. \square

مثال ۱-۸-۴. برای گراف شکل ۱-۲، درجه (D) ، و لاپلاسیان (L) محاسبه شده است. همانطور که ملاحظه می شود، ماتریس لاپلاسیان متقارن است.

$$W = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix}, D = \begin{pmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}, L = \begin{pmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 & 0 \\ -1 & -1 & 4 & -1 & -1 \\ -1 & -1 & -1 & 4 & -1 \\ -1 & 0 & -1 & -1 & 3 \end{pmatrix}$$

مثال ۱-۸-۵. با تجزیه ویژه ماتریس L گراف مثال قبل، مقادیر ویژه آن برابر است با:

$$\lambda = \{0, 3, 5, 5, 5\}$$

ماتریس L دارای یک مقدار ویژه صفر است و گراف مربوطه نیز همبند است بنابراین همانطور که در قضیه ۱-۸-۳ بیان شد تعداد اجزاء همبند گراف برابر است با تکرار مقدار ویژه صفر ماتریس لاپلاسیان گراف.

۹-۱ افزایش گراف

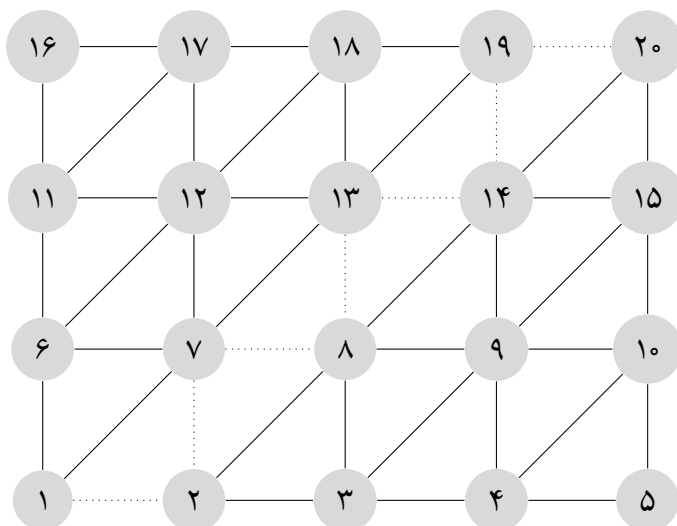
هدف مسئله افزایش k تایی گراف، حذف یال‌های گراف به نحوی است که باعث ایجاد k زیرگراف مجزا شود [۱۱]. گراف $\tilde{G} = (V, E, W_V, W_E)$ را در نظر می‌گیریم، رأس v_i در V نماینده یک رأس گراف و یال $e = (i, j)$ در E متصل‌کننده دو رأس i و j و W_V و W_E مجموعه‌هایی هستند که به ترتیب نشان‌دهنده وزنهای نامنفی برای هر رأس و هر یال هستند. در این صورت $W_V v_i$ وزن رأس i و $W_E e_{ij}$ وزن یال e_{ij} را نشان می‌دهد. افزایش G به معنی تقسیم کردن V به اجتماعی از n مجموعه مجزا V_1, V_2, \dots, V_n است طوری که

$$V = V_1 \cup V_2 \cup \dots \cup V_n \quad (6-1)$$

برای افزایش بهینه شرایط زیر باید برقرار باشند [۱۱]:

۱. مجموع وزن‌های W_V رأس‌ها در هر V_i باید تقریباً برابر باشند.
 ۲. مجموع وزن‌های W_E از یال‌های اتصال‌دهنده رأس‌ها در V_i و V_j متفاوت باید مینیمم باشد.
- در ادامه وزن هر رأس و هر یال را یک در نظر می‌گیریم.

مثال ۱-۹-۱. شکل ۱-۳ یک گراف افزایش شده را نشان می‌دهد. رأس‌های $\{1, 6, 7, 11, 12, 13, 16, 17, 18, 19\}$ متعلق به



شکل ۱-۳: یک گراف افراز شده

زیرگراف V_1 و رأس‌های $\{2, 3, 4, 5, 8, 9, 10, 14, 15, 20\}$ متعلق به زیرگراف V_2 و یالهایی که به صورت نقطه‌چین مشخص شده‌اند یال‌های برش هستند. همانطور که مشاهده می‌کنید هر دو خوشه به دست آمده دارای اندازه یکسان‌اند و اندازه برش یالی خوشه‌های V_1 و V_2 هفت است.

تعریف ۱-۹-۲. بردارهای غیر صفر u و v را نسبت به ماتریس متقارن و معین مثبت A مزدوج گوئیم اگر

$$u^T A v = 0$$

از آنجایی که A متقارن و معین مثبت است، سمت چپ معادله بالا یک ضرب داخلی است.

$$\langle u, v \rangle_A := \langle Au, v \rangle = \langle u, A^T v \rangle = \langle u, Av \rangle = u^T Av.$$

دو بردار مزدوج‌اند اگر تنها اگر نسبت به ضرب داخلی متعامد باشند. مزدوج بودن یک خاصیت متقارن است، اگر u با v مزدوج باشد آنگاه، v نیز با u مزدوج است. اگر $P = \{p_1, \dots, p_n\}$ مجموعه‌ای از n بردار متقابلاً مزدوج باشد آنگاه، P یک پایه برای \mathbb{R}^n تشکیل می‌دهد.

اگر u و v دو بردار متناظر با مقدار ویژه صفر باشند مزدوج‌اند اما متعامد نیستند. پس u و v باید متناظر با مقدار ویژه غیر صفر باشد.

۱۰-۱ الگوریتم k-means برای خوشه‌بندی داده‌ها

k-means یکی از ساده‌ترین الگوریتم‌های یادگیری بدون ناظر است که مسئله خوشه‌بندی را حل می‌کند و راه‌حلی آسان را برای کلاس‌بندی یک مجموعه داده به تعداد مشخصی از خوشه‌ها (به طور مثال k خوشه) ارائه می‌دهد. ایده اصلی تعریف k مرکز برای هر خوشه است. به دلیل این که مکان‌های مختلف مراکز، نتایج متفاوتی در خوشه‌بندی دارند این مراکز باید به شیوه هوشیارانه‌ای مکان‌دهی اولیه شوند. می‌توان آن‌ها را به گونه‌ای انتخاب کرد که تا جای ممکن دور از یکدیگر قرار داشته باشند. گام بعدی این است که هر نقطه متعلق به مجموعه داده را به نزدیک‌ترین مرکز اختصاص دهیم. با اختصاص‌دهی تمام نقاط به مراکز، مرحله اول تکمیل و گروه‌بندی اولیه انجام می‌شود. در مرحله بعدی باید k مرکز جدید برای خوشه‌های گام قبلی مجدداً محاسبه شوند. پس از محاسبه این k مرکز جدید باید یک اتصال جدید بین نقاط مجموعه داده و نزدیک‌ترین مرکز جدید انجام شود. یک حلقه ایجاد می‌شود. در نتیجه این حلقه، k مرکز گام به گام مکان خود را تغییر می‌دهند تا زمانی که تغییری انجام نشود. به عبارت دیگر، مراکز دیگر حرکت نکنند [۱۳].

در نهایت، این الگوریتم یک تابع هدف را کمینه می‌کند که در این مورد تابع مربع خطا است. تابع هدف به صورت

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

تعریف می‌شود که در آن $\|x_i^{(j)} - c_j\|^2$ مقیاس اندازه‌گیری فاصله انتخاب شده بین داده $x_i^{(j)}$ و مرکز خوشه c_j است که شاخصی برای فاصله n داده از مرکز خوشه‌های مربوطه است.

اگر چه این روش همیشه متوقف خواهد شد [۱۴]، اما الگوریتم k-means لزوماً پیکربندی مطلوب را که کمینه سراسری تابع هدف است، پیدا نمی‌کند. همچنین الگوریتم به مراکز اولیه خوشه‌ها که به صورت تصادفی انتخاب شده‌اند حساس است. می‌توان الگوریتم را چندین بار اجرا کرد تا این اثر کاهش یابد.

k-means یک الگوریتم ساده است که برای بسیاری از مسائل مورد استفاده قرار می‌گیرد. فرض کنید که n نقطه x_1, x_2, \dots, x_n داریم که همگی عضو یک کلاس هستند و می‌دانیم که به $k < n$ خوشه تقسیم می‌شوند. فرض کنید m_i میانگین نقاط خوشه i باشد. اگر خوشه‌ها به خوبی تفکیک شده باشند، می‌توانیم از طبقه‌بند کمترین فاصله استفاده کنیم تا آن‌ها را جدا کنیم. یعنی می‌توان گفت که x جدید در خوشه i است اگر $\|x - m_i\|$ کمتر از تمام k فواصل دیگر باشد. می‌توان k میانگین را از الگوریتم ۱-۱ به دست آورد.

مثال ۱-۱۰-۱. n نقطه داده را همان طور که در شکل ۱-۴ (آ) نمایش داده شده است در نظر بگیرید. می‌خواهیم این نقاط را به سه خوشه تقسیم کنیم. ابتدا سه نقطه از نقاط را به عنوان مراکز خوشه‌ها به صورت تصادفی انتخاب می‌کنیم. در شکل ۱-۴ (آ) این نقاط به رنگ‌های قرمز، آبی و سبز به نمایش درآمده‌اند. فاصله تمام نقاط دیگر را با این مراکز محاسبه می‌کنیم و هر نقطه را به مرکزی که کمترین فاصله از آن را دارد اختصاص می‌دهیم. خوشه‌بندی حاصل از این مراکز اولیه در شکل ۱-۴ (ب) نمایش داده شده است و نقاط به رنگ خوشه‌ای که به آن اختصاص یافته‌اند، رنگ‌آمیزی شده‌اند. در مرحله بعد،

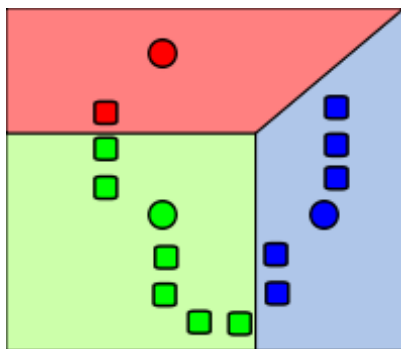
الگوریتم ۱-۱ الگوریتم k-means برای یافتن k میانگین

Require: n sample data points x_1, x_2, \dots, x_n

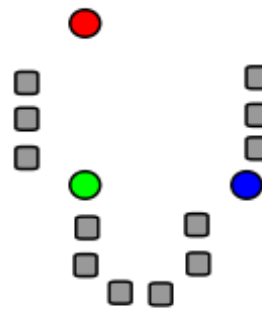
Ensure: k means.

- 1: Make initial guesses for the means m_1, m_2, \dots, m_k
- 2: **Until** there are no changes in any mean
- 3: Use the estimated means to classify the samples into clusters
- 4: **for** $i = 1$ to k **do**
- 5: Replace m_i with the mean of all of the samples for cluster i
- 6: **end for**
- 7: **end Until**

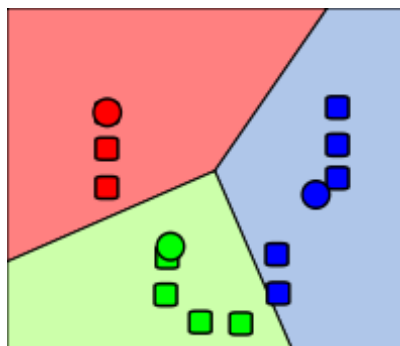
میانگین هر کدام از خوشه‌ها را محاسبه می‌کنیم و مراکز خوشه‌ها را به این میانگین‌ها انتقال می‌دهیم. در شکل ۱-۴(ج) این انتقال با فلش نشان داده شده است. مراحل قبلی را برای این مراکز جدید تکرار می‌کنیم تا در نهایت مراکز دیگر تغییری نکنند. شکل ۱-۴(د) خوشه‌بندی نهایی را برای این نقاط نشان می‌دهد.



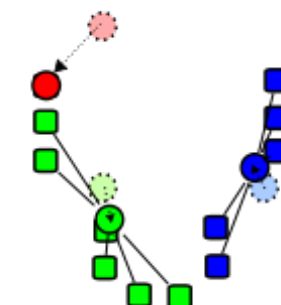
(ب) k خوشه تولید شده توسط میانگین‌های اولیه



(ا) k میانگین تصادفی اولیه



(د) خوشه‌بندی نهایی الگوریتم k-means



(ج) تغییر مراکز خوشه‌ها به میانگین‌های جدید

شکل ۱-۴: مراحل روش k-means برای خوشه‌بندی داده‌ها

فصل ۲

روش‌های افرازبندی طیفی گراف

روش افرازبندی طیفی گراف یک روش دقیق و کارآمد است. در این روش نیاز به محاسبه بردارهای ویژه ماتریس لاپلاسیان متناظر با گراف است که برای گراف‌های متناظر با مجموعه داده‌های بزرگ هزینه محاسباتی و حافظه‌ای بالایی دارد. در این فصل ابتدا روش افرازبندی طیفی گراف را شرح داده و سپس به بیان دو روش تقریبی گرادیان مزدوج و نیستروم برای محاسبه بردارهای ویژه ماتریس لاپلاسیان می‌پردازیم تا به کمک آن‌ها روش افرازبندی طیفی را بر روی مجموعه داده‌های بزرگ پیاده‌سازی کنیم.

۱-۲ افرازبندی طیفی گراف

گراف بدون جهت $G = (V, E)$ را وقتی که V مجموعه رئوس و E مجموعه یال‌های آن باشد، در نظر بگیرید. هدف افرازبندی طیفی گراف (در حالت ساده به دو بخش)، افراز رئوس گراف به دو مجموعه است، به طوری که تعداد یال‌های برشی گراف کمینه شود و همچنین تعداد رأس‌های دو مجموعه با هم برابر باشند. فرض کنید $|V| = n$ باشد و x یک جواب شدنی برای دو بخشی کردن طیفی گراف باشد و مجموعه‌های افراز V_1 و V_2 را تولید کند. در این صورت می‌توان x را به شکل بردار n تایی با مولفه‌های ۱ و -۱ در نظر گرفت. مجموعه‌های افراز V_1 و V_2 به صورت

$$V_1 = \{i : x_i = 1\}, \quad V_2 = \{i : x_i = -1\} \quad (1-2)$$

تعریف می‌شوند که در آن x_i مولفه i ام بردار x است. از طرفی باید تعداد رأس‌ها در یک مجموعه افراز، با تعداد رأس‌ها در مجموعه دیگر برابر باشند. در نتیجه، جواب شدنی x در روش دو بخشی کردن طیفی گراف، باید در شرط

$$e^T x = 0 \quad (2-2)$$

صدق کند که در آن، e برداری n تایی است که همه مولفه‌های آن برابر یک هستند. با استفاده از لم ۱-۲-۱، تعداد یال‌های بین دو مجموعه افراز محاسبه می‌شود.

لم ۱-۲-۱. گراف $G = (V, E)$ را در نظر بگیرید. فرض کنید x جواب شدنی برای روش دو بخشی کردن طیفی گراف باشد و مجموعه‌های V_1 و V_2 را به فرم ۱-۲ تولید کند. همچنین، فرض کنید L ماتریس لاپلاسیان متناظر با گراف G باشد. در این صورت، تعداد یال‌های بین دو مجموعه افراز که توسط x تولید شده‌اند، برابر با

$$\frac{1}{4} (x^T L x) \quad (3-2)$$

می‌باشد.

برهان. فرض کنید x جواب شدنی برای روش دو بخشی کردن طیفی گراف باشد. در این صورت اگر دو راس i و j در گراف G ، نسبت به جواب شدنی x ، در یک مجموعه افراز قرار داشته باشند می‌گوییم $i \sim j$. همچنین اگر i و j در یک مجموعه افراز نباشند، می‌گوییم $i \not\sim j$.

برای جواب شدنی x رابطه $\frac{1}{4} x^T L x$ را در نظر بگیرید. در این صورت داریم:

$$\frac{1}{4} x^T L x = \frac{1}{4} x^T (D - A) x = \frac{1}{4} \left(\sum_{i=1}^n d(i) x_i^2 - \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \right) \quad (4-2)$$

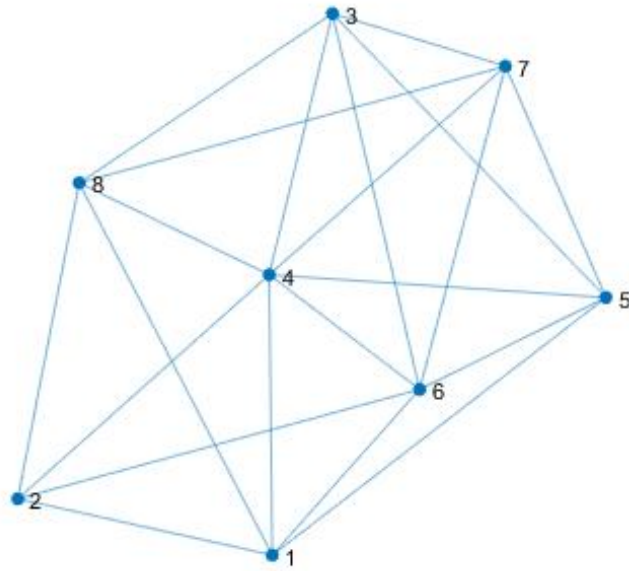
از آنجایی که برای هر یال $e \in E$ دو راس $i, j \in V$ وجود دارند به طوری که $(i, j) = e \in E$ ، پس رابطه (۴-۲) به فرم

$$\begin{aligned} \frac{1}{4} x^T L x &= \frac{1}{4} \left(\sum_{k=1}^{|E|} 2 - \sum_{(i,j) \in E: i \sim j} 2 + \sum_{(i,j) \in E: i \not\sim j} 2 \right) \\ &= \frac{1}{4} \left(\sum_{(i,j) \in E: i \sim j} 2 + \sum_{(i,j) \in E: i \not\sim j} 2 - \sum_{(i,j) \in E: i \sim j} 2 + \sum_{(i,j) \in E: i \not\sim j} 2 \right) \\ &= \frac{1}{4} \left(\sum_{(i,j) \in E: i \not\sim j} 4 \right) \\ &= \sum_{(i,j) \in E: i \not\sim j} 1 \end{aligned}$$

نتیجه می‌شود. بنابراین مشاهده می‌شود که تعداد یال‌های برشی، که توسط جواب شدنی x تولید شده‌اند، از حاصل ضرب $\frac{1}{4} x^T L x$ محاسبه می‌شود. \square

در روش طیفی برای افراز بندی گراف، از بردار ویژه متناظر با دومین کوچکترین مقدار ویژه به عنوان جواب مسئله استفاده می‌شود. در ادامه این روش را با ذکر یک مثال توضیح می‌دهیم.

مثال ۲-۲



شکل ۱-۲: گراف با ۸ رأس

گراف شکل ۱-۲ را در نظر بگیرید. ماتریس مجاورت این گراف به صورت

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

است و وزن یال‌ها برابر با یک در نظر گرفته شده‌اند. ماتریس درجه این گراف نیز برابر با

$$D = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 7 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix}$$

می‌باشد. با استفاده از فرمول $L = D - W$ ماتریس لاپلاسیان گراف را محاسبه می‌کنیم:

$$L = \begin{bmatrix} 5 & -1 & 0 & -1 & -1 & -1 & 0 & -1 \\ -1 & 4 & 0 & -1 & 0 & -1 & 0 & -1 \\ 0 & 0 & 5 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 7 & -1 & -1 & -1 & -1 \\ -1 & 0 & -1 & -1 & 5 & -1 & -1 & 0 \\ -1 & -1 & -1 & -1 & -1 & 6 & -1 & 0 \\ 0 & 0 & -1 & -1 & -1 & -1 & 5 & -1 \\ -1 & -1 & -1 & -1 & 0 & 0 & -1 & 5 \end{bmatrix}$$

مقادیر ویژه ماتریس لاپلاسیان و بردارهای ویژه متناظر با آن‌ها را محاسبه می‌کنیم:

$$\lambda = \begin{bmatrix} -0,00003/4228 \\ 4,7080 \\ 5,4781 \\ 6,0000 \\ 6,6677 \\ 7,7235 \\ 8,0000 \end{bmatrix}$$

$$S = \begin{bmatrix} 0,3536 & 0,2998 & 0,3899 & 0,5948 & -0,0000 & 0,3420 & 0,3801 & 0,1336 \\ 0,3536 & 0,6960 & -0,0645 & -0,5138 & -0,0000 & -0,2658 & 0,1844 & 0,1336 \\ 0,3536 & -0,3864 & -0,2519 & -0,1552 & -0,7071 & 0,1142 & 0,3275 & 0,1336 \\ 0,3536 & -0,0000 & -0,0000 & -0,0000 & 0,0000 & 0,0000 & 0,0000 & -0,9354 \\ 0,3536 & -0,3249 & 0,5226 & 0,0648 & -0,0000 & -0,6715 & -0,1528 & 0,1336 \\ 0,3536 & -0,0395 & 0,2664 & -0,3154 & 0,0000 & 0,5496 & -0,6189 & 0,1336 \\ 0,3536 & -0,3864 & -0,2519 & -0,1552 & 0,7071 & 0,1142 & 0,3275 & 0,1336 \\ 0,3536 & 0,1414 & -0,6107 & 0,4800 & -0,0000 & -0,1826 & -0,4478 & 0,1336 \end{bmatrix}$$

همانگونه که در فصل اول اشاره شد، تکرر جبری مقدار ویژه صفر، برابر با مولفه‌های همبندی گراف است که در اینجا چون یک گراف همبند داریم، برابر با یک است. طبق قضیه ۱-۳-۲۳ دومین مقدار کمینه عبارت $\frac{x^T L x}{x^T x}$ برابر با دومین کوچکترین مقدار ویژه ماتریس لاپلاسیان می‌باشد و بردار ویژه متناظر با این مقدار ویژه جواب این عبارت است. بنابراین برای افراز گراف به دو زیر گراف، از بردار ویژه متناظر با دومین کوچکترین مقدار ویژه استفاده می‌کنیم یک روش استفاده از این بردار ویژه به این ترتیب است که میانه مولفه‌های این بردار را می‌یابیم و هر مولفه که کمتر یا مساوی با این مقدار بود، راس متناظر با آن را به زیرگراف اول و مولفه‌هایی که مقدارشان بیشتر از میانه بود را به زیر گراف دوم اختصاص می‌دهیم. برای بردار ویژه متناظر با دومین کوچکترین مقدار ویژه ماتریس لاپلاسیان این گراف داریم:

$$\text{median } s_2 = -0,0198$$

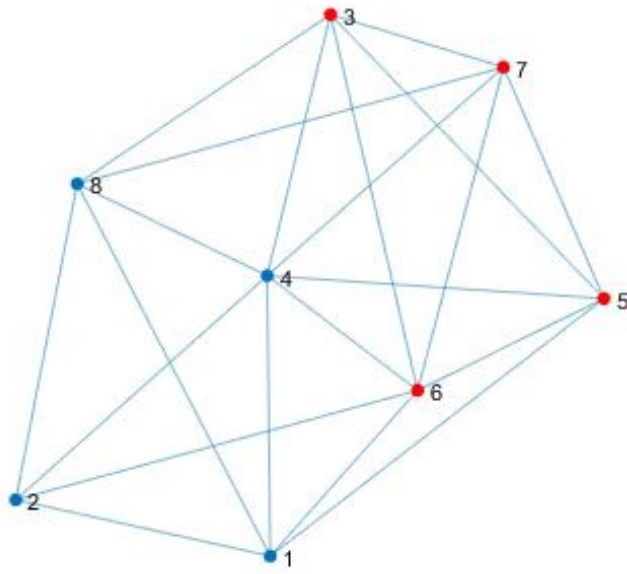
مولفه‌های دوم، پنجم و ششم این بردار کمتر مساوی میانه بردار بوده پس رئوس متناظر با آن‌ها یعنی راس‌های دوم، پنجم و ششم را به زیر گراف اول اختصاص می‌دهیم و $V_1 = \{3, 5, 6, 7\}$ ، به همین منوال مولفه‌های اول، سوم و چهارم بیشتر از ۰ بوده و به زیرگراف دوم اختصاص داده می‌شوند و $V_2 = \{1, 2, 4, 8\}$. زیرگراف‌های حاصل از این افراز در شکل ۲-۲ نشان داده شده‌اند. راس‌های زیرگراف V_1 با رنگ قرمز نمایش داده شده‌اند. بردار x حاصل از این روش برابر با

$$x' = \begin{bmatrix} -1 & -1 & 1 & -1 & 1 & 1 & 1 & -1 \end{bmatrix}$$

می‌باشد. با قرار دادن این بردار در معادله (۲-۳) نتیجه زیر حاصل می‌شود:

$$\text{تعداد یال‌های برشی} = \frac{1}{4} x^T L x = 9$$

همانگونه که مشاهده می‌شود $|V_1| = |V_2| = 4$ و تعداد یال‌های برشی کمینه است.



شکل ۲-۲: افراز حاصل از روش طیفی

Require: a graph $G = (V, E)$

Ensure: graphs $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$

- 1: compute the eigenvector s_2 corresponding to the second-smallest eigenvalue
 - 2: search median of s_2
 - 3: **for** $i = 1$ to $|V|$ **do**
 - 4: **if** $s_2(i) \leq \text{median}$ **then**
 - 5: put node i in partition V_1
 - 6: **else**
 - 7: put node i in partition V_2
 - 8: **end if**
 - 9: **end for**
 - 10: **if** $|V_1| - |V_2| > 1$ **then**
 - 11: move some vertices with components equal to median from V_1 to V_2 to make this difference at most one
 - 12: **end if**
 - 13: let V'_1 be the set of vertices in V_1 adjacent to some vertex in V_2
let V'_2 be the set of vertices in V_2 adjacent to some vertex in V_1
set up the edge separator E' - the set of edges of G with one point in V'_1 and the second in V'_2
 - 14: let E_1 be the set of edges with both end vertices in V_1
let E_2 be the set of edges with both end vertices in V_2
set up the graphs $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$
-

۳-۲ برش k تایی همزمان با چند بردار ویژه

برای انجام یک خوشه‌بندی k تایی می‌توان این کار را به دو روش بازگشتی و چندگانه انجام داد.

در خوشه‌بندی طیفی چندگانه از k بردار ویژه ماتریس لاپلاسیان به طور همزمان برای افراز گراف استفاده می‌شود [۱۵] که افراز k تایی نامیده می‌شود. در این روش ابتدا بردارهای ویژه متناظر با k تا از کوچکترین مقادیر ویژه انتخاب شده و خوشه‌بندی یا براساس علامت داده‌ها در بردارهای ویژه دوم تا k ام انجام می‌گیرد که بر این اساس، داده‌هایی که تغییر علامت روی این k بردار برای آنها یکسان باشد، در یک گروه قرار خواهند گرفت.

در روش بازگشتی، بعد از تقسیم داده‌ها به دو خوشه بررسی می‌شود که آیا تعداد خوشه‌ها متناسب با نیاز است یا خیر. اگر لازم باشد، همین فرآیند مجدداً بر روی زیر خوشه‌های بدست آمده اعمال می‌شود و تا رسیدن به تعداد خوشه مطلوب کار ادامه می‌یابد [۱۶].

۴-۲ فرمول‌بندی روش طیفی برای بکارگیری روش‌های بهینه‌سازی

مسئله دو بخشی کردن گراف با روش طیفی، با استفاده از لم ۱-۱-۲ و روابط (۱-۲) و (۲-۲) به فرم

$$\begin{aligned} \min \quad & x^T Lx \\ \text{s.t.} \quad & e^T x = 0 \\ & x_i \in \{-1, +1\} \quad \forall 1 \leq i \leq n \end{aligned} \quad (۵-۲)$$

فرمول‌بندی می‌شود. مسئله (۵-۲)، یک مسئله بهینه‌سازی عدد صحیح است. حل این مسئله در حالتی که تعداد راس‌های گراف G عددی بزرگ باشد، مشکل خواهد بود. بر این اساس، با آزادسازی^۱ قید عدد صحیح بودن x ، مسئله

$$\begin{aligned} \min \quad & x^T Lx \\ \text{s.t.} \quad & e^T x = 0 \end{aligned} \quad (۶-۲)$$

نتیجه می‌شود. بنا بر لم ۱-۱-۲، به سادگی نتیجه می‌شود که جواب بهینه مسئله (۶-۲)، بردار صفر است. برای حل این مشکل باید قید دیگری به مسئله (۶-۲) اضافه کرد به طوری که جواب‌های شدنی مسئله (۵-۲) برای مسئله (۶-۲) هم شدنی باشند، و جواب بهینه مسئله (۶-۲) بردار صفر نباشد. در نتیجه مسئله بهینه‌سازی به فرم

$$\begin{aligned} \min \quad & x^T Lx \\ \text{s.t.} \quad & e^T x = 0 \\ & \sum_{i=1}^n x_i^2 = n. \end{aligned} \quad (۷-۲)$$

به دست می‌آید [۱۷]. ساختار مسئله (۷-۲) نشان می‌دهد جواب‌های شدنی مسئله (۵-۲) برای مسئله (۷-۲) هم شدنی هستند، اما ممکن است جواب بهینه مسئله (۷-۲) یک جواب شدنی برای مسئله (۵-۲) نباشد. به مثال زیر توجه کنید.

^۱Relaxation

مثال ۲-۴-۱. فرض کنید ماتریس لاپلاسیس متناظر با گراف بدون جهت G ، به شکل

$$L = \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

باشد. نمونه‌ای از مسئله (۵-۲) را به صورت

$$\begin{aligned} \min \quad & (x_1 - x_2)^2 + (x_1 - x_3)^2 \\ \text{s.t.} \quad & x_1 + x_2 + x_3 + x_4 = 0 \\ & x_i \in \{-1, +1\} \quad \forall 1 \leq i \leq 4 \end{aligned}$$

در نظر بگیرید. مجموعه جواب این نمونه به صورت

$$S = \{(-1, -1, 1, 1), (-1, 1, 1, -1), (-1, 1, -1, 1), (1, -1, -1, 1), (1, -1, 1, -1), (1, 1, -1, -1)\}$$

به دست می‌آید. کمترین مقدار تابع هدف در این مجموعه جواب برابر با ۴ است. حال، نمونه‌ای دیگر از مسئله (۷-۲) را به صورت

$$\begin{aligned} \min \quad & (x_1 - x_2)^2 + (x_1 - x_3)^2 \\ \text{s.t.} \quad & x_1 + x_2 + x_3 + x_4 = 0 \\ & x_1^2 + x_2^2 + x_3^2 + x_4^2 = 4 \end{aligned}$$

در نظر بگیرید. بردار $(0, 0, -\sqrt{2}, +\sqrt{2})$ یک جواب شدنی برای این نمونه است و مقدار تابع هدف در این بردار برابر با ۲ است. در نتیجه مشاهده می‌کنیم که جواب بهینه مسئله (۷-۲) ممکن است جواب شدنی برای مسئله (۵-۲) نباشد.

از مثال ۲-۴-۱ و ساختار مسئله (۷-۲) مشاهده می‌شود که جواب بهینه مسئله (۷-۲)، یک کران پایین برای جواب بهینه مسئله (۵-۲) است. از آنجایی که در روش دو بخشی کردن طیفی گراف، هدف یافتن یک جواب تقریبی قابل قبول برای مسئله دو بخشی کردن گراف است می‌توان به جای حل مسئله (۵-۲) مسئله (۷-۲) را حل نمود.

پوشن^۱ و همکارانش در سال ۱۹۹۰ نشان دادند، بردار ویژه متناظر با دومین - کوچکترین مقدار ویژه ماتریس لاپلاسی L ، جواب بهینه مسئله (۷-۲) است. هسیه^۲ و همکارانش در سال ۱۹۹۵ روشی ارائه دادند که براساس آن، اگر x^* جواب بهینه مسئله (۷-۲) باشد، با استفاده از یک بردار میانگین، x^* را به یک جواب شدنی برای مسئله (۵-۲)، با مجموعه‌های افراز به شکل (۱-۲) تبدیل می‌کند.

۵-۲ مسئله بهینه‌سازی نامقید، معادل با مسئله (۷-۲)

همان‌گونه که در بخش ۴-۲ نشان داده شد، بردار x ، اگر در رابطه

$$Lx = \lambda x \quad (۸-۲)$$

صدق کند جواب بهینه مسئله (۷-۲) است

$$Lx = \lambda_2 x \quad (۹-۲)$$

که در آن λ_2 دومین - کوچکترین مقدار ویژه ماتریس لاپلاسی L است. همچنین بنا به تعریف بردار ویژه به جای رابطه (۹-۲) می‌توان رابطه

$$\lambda = \frac{x^T Lx}{x^T x} \quad (۱۰-۲)$$

را در نظر گرفت. بر این اساس، به جای حل مسئله (۷-۲)، کافی است بردار x را طوری بیابیم که در رابطه (۱۰-۲) صدق کند. برای این منظور کافی است کمینه تابع

$$F(x) = \frac{x^T Lx}{x^T x} \quad (۱۱-۲)$$

محاسبه شود. یکی از روش‌های یافتن کمینه تابع F ، روش گرادیان مزدوج است که در بخش ۸-۲ به بیان آن می‌پردازیم.

^۱Pothen ^۲Hsie

۶-۲ روش گرادیان نزولی

استفاده از منفی بردار گرادیان به عنوان یک جهت برای کمینه‌سازی، اولین بار توسط کوشی^۱ در سال ۱۸۴۷ صورت گرفت. در این روش، از یک نقطه آزمون اولیه X_i شروع و به طور تکراری طبق قاعده

$$X_{i+1} = X_i + \lambda_i^* S_i \quad (۱۲-۲)$$

به سمت نقطه بهینه حرکت می‌کنیم که λ_i^* طول گام بهینه در امتداد جهت جستجوی $S_i = -\nabla f_i$ است. برای یافتن طول گام بهینه λ_i^*

$$f(X_i + \lambda_i S_i)$$

را نسبت به λ_i کمینه می‌کنیم.

۱-۶-۲ معیار همگرایی

معیارهای زیر را می‌توان به عنوان شرایط توقف الگوریتم لحاظ کرد:

$$1. \quad \left| \frac{f(X_{i+1}) - f(X_i)}{f(X_i)} \right| \leq \varepsilon_1$$

$$2. \quad \left| \frac{\partial f}{\partial x_i} \right| \leq \varepsilon_2, i = 1, 2, \dots, n$$

$$3. \quad |X_{i+1} - X_i| \leq \varepsilon_3$$

روش گرادیان نزولی، از آنجا که در هر جستجوی یک بعدی از بهترین جهت شروع می‌شود، ممکن است بهترین روش کمینه‌سازی نامقید به نظر برسد. اما به دلیل این که جهت گرادیان نزولی یک خاصیت موضعی است، این روش در بیشتر مسائل کارا نیست.

۷-۲ روش گرادیان مزدوج

مشخصات روش گرادیان نزولی را می‌توان با تعدیل آن به یک روش گرادیان مزدوج (فلچر-ریویس^۲) تا حد زیادی بهبود بخشید. نشان داده شده است که هر روش کمینه‌سازی که از جهات مزدوج استفاده کند، به لحاظ درجه دو همگراست [۱۲]. این خاصیت همگرایی درجه دوم بسیار مفید است. زیرا که کمینه شدن یک تابع درجه دوم را در تعداد n گام یا کمتر تضمین می‌کند.

^۱Cauchy

^۲Fletcher-Reeves

از آنجا که هر تابع عمومی را در نزدیکی نقطه بهینه می‌توان با یک تابع درجه دوم تقریب زد، انتظار می‌رود هر روشی که به لحاظ درجه دوم همگرا باشد، نقطه بهینه را در تعداد تکرارهای محدودی پیدا کند. روش گرادیان نزولی را می‌توان با تعدیل آن به یک روش گرادیان مزدوج تا حد زیادی بهبود بخشید. این روش هر جهت جستجوی جدید را به صورت ترکیبی خطی از همه جهات جستجوی قبلی و گرادیان‌های جدید تعیین شده است، به دست می‌آورد [۱۲].
قضیه زیر در بسط روش گرادیان مزدوج دارای اهمیت است.

قضیه ۲-۷-۱. فرض کنید که پس از i گام کمینه‌سازی تابع

$$f(X) = \frac{1}{2} X^T A X + B^T X + C$$

به نقطه X_{i+1} رسیده‌ایم. اگر جهات جستجوی مورد استفاده در فرآیند کمینه‌سازی یعنی S_1, S_2, \dots, S_i نسبت به A متقابلاً مزدوج باشند، آنگاه:

$$S_k^T \nabla f_{i+1} = 0, \quad k = 1, 2, \dots, i$$

برهان. گرادیان تابع f در X_{i+1} به صورت

$$\nabla f_{i+1} = A X_{i+1} + B \quad (13-2)$$

می‌باشد. مقدار X_{i+1} پس از i گام کمینه‌سازی به صورت

$$\begin{aligned} X_{i+1} &= X_1 + \lambda_1^* S_1 + \lambda_2^* S_2 + \dots + \lambda_k^* S_k + \lambda_{k+1}^* S_{k+1} + \dots + \lambda_i^* S_i \\ &= X_{k+1} + \sum_{j=k+1}^i \lambda_j^* S_j \end{aligned} \quad (14-2)$$

حاصل می‌شود که λ_j^* طول گام بهینه‌سازی در جهت S_j است.
با توجه به رابطه (۱۳-۲)، رابطه (۱۴-۲) به صورت

$$\begin{aligned} \nabla f_{i+1} &= A \left(X_{k+1} + \sum_{j=k+1}^i \lambda_j^* S_j \right) + B \\ &= \nabla f_{k+1} + \sum_{j=k+1}^i \lambda_j^* A S_j \end{aligned} \quad (15-2)$$

به دست می‌آید. با ضرب S_k^T در هر دو طرف رابطه (۲-۱۵)

$$S_k^T \nabla f_{i+1} = S_k^T \nabla f_{k+1} + \sum_{j=k+1}^i \lambda_j^* S_k^T A S_j \quad (۲-۱۶)$$

نتیجه می‌شود. از آنجا که λ_k^* ، طول گام کمینه‌سازی، در امتداد جهت S_k است، عبارت اول سمت راست رابطه (۲-۱۶) صفر است، و چون جهات جستجوی S_1, S_2, \dots, S_i ، ماتریس A مزدوج هستند، عبارت دوم هم صفر است. پس نتیجه مطلوب به دست می‌آید. در نتیجه

$$S_k^T \nabla f_{i+1} = 0, \quad k = 1, 2, \dots, i \quad (۲-۱۷)$$

□

۱-۷-۲ الگوریتم فلچر-ریویس

در این بخش بسط الگوریتم فلچر-ریویس بیان می‌شود. در این الگوریتم روش تندترین کاهش را که بر یک تابع درجه دوم $f(X) = \frac{1}{2} X^T A X + B^T X + C$ اعمال شده است، با تحمیل شرط متقابلاً مزدوج بودن جهات متوالی تعدیل می‌کنیم. فرض کنید X_1 ، نقطه آغازین برای کمینه‌سازی، و اولین جهت جستجو، جهت تندترین کاهش باشد. آنگاه

$$S_1 = -\nabla f_1 = -A X_1 - B \quad (۲-۱۸)$$

$$X_2 = X_1 + \lambda_1^* S_1 \quad (۲-۱۹)$$

که λ_1^* ، طول گام بهینه‌سازی در جهت S_1 به گونه‌ای است که

$$S_1^T \nabla f|_{X_2} = 0 \quad (۲-۲۰)$$

با بسط این رابطه به صورت

$$S_1^T (A(X_1 + \lambda_1^* S_1) + B) = 0$$

$$S_1^T A X_1 + \lambda_1^* S_1^T A S_1 + S_1^T B = 0$$

از این رابطه مقدار λ_1^* به صورت

$$\lambda_1^* = \frac{-S_1^T (AX_1 + B)}{S_1^T AS_1} = \frac{S_1^T \nabla f_1}{S_1^T AS_1} \quad (21-2)$$

به دست می آید. حال دومین جهت جستجو را به عنوان یک ترکیب خطی از S_1 و $-\nabla f_2$ به صورت

$$S_2 = -\nabla f_2 + \beta_2 S_1 \quad (22-2)$$

بیان می کنیم. لازم به ذکر است که β_2 باید به گونه ای انتخاب شود که S_1 و S_2 مزدوج شوند. در نتیجه

$$S_1^T AS_2 = 0 \quad (23-2)$$

با قرار دادن مقدار S_2 از رابطه (22-2)، رابطه (23-2) به صورت

$$S_1^T A(-\nabla f_2 + \beta_2 S_1) = 0 \quad (24-2)$$

حاصل می شود. از رابطه (19-2) مقدار S_1 به صورت

$$S_1 = \frac{(X_2 - X_1)}{\lambda_1^*} \quad (25-2)$$

حاصل می شود. در نتیجه رابطه (24-2) به صورت

$$S_1^T AS_2 = -\frac{(X_2 - X_1)}{\lambda_1^*} A(\nabla f_2 - \beta_2 S_1) = 0 \quad (26-2)$$

به دست می آید. تفاضل گرادیان ها به صورت

$$(\nabla f_2 - \nabla f_1) = (AX_2 + B) - (AX_1 + B) = A(X_2 - X_1) \quad (27-2)$$

تعریف می شود. با توجه به مقارن بودن ماتریس A و به کمک رابطه (27-2)، رابطه (26-2) به فرم

$$(\nabla f_2 - \nabla f_1)^T (\nabla f_2 - \beta_2 S_1) = 0 \quad (28-2)$$

به دست می‌آید. بسط رابطه (۲۸-۲) به صورت

$$\nabla f_2^T \nabla f_2 - \nabla f_1^T \nabla f_2 - \beta_2 \nabla f_2^T S_1 + \beta_2 \nabla f_1^T S_1 = 0$$

به دست می‌آید. با استفاده از رابطه (۱۷-۲) داریم:

$$\nabla f_1^T \nabla f_2 = -S_1^T \nabla f_2 = 0$$

در نتیجه مقدار β_2 به صورت

$$\beta_2 = \frac{\nabla f_2^T \nabla f_2}{\nabla f_1^T \nabla f_1}$$

به دست می‌آید. اگر جهت جستجو را ترکیب خطی از S_1 و $S_2 - \nabla f_3$ در نظر بگیریم مقدار S_3 به صورت

$$S_3 = -\nabla f_3 + \beta_3 S_2 + \delta_3 S_1 \quad (29-2)$$

به دست می‌آید که مقادیر β_3 ، δ_3 از مزدوج قرار دادن S_3 با S_1 و S_2 نتیجه می‌شود.

رابطه

$$S_1^T A S_3 = -S_1^T A \nabla f_3 + \beta_3 S_1^T A S_2 + \delta_3 S_1^T A S_1 = 0 \quad (30-2)$$

را در نظر بگیرید. اگر S_1 و S_2 مزدوج در نظر گرفته شوند داریم:

$$S_1^T A S_2 = 0$$

با استفاده از رابطه (۳۰-۲)

$$\delta_3 = \frac{S_1^T A \nabla f_3}{S_1^T A S_1} \quad (31-2)$$

نتیجه می‌شود. از رابطه (۲۵-۲) مقدار δ_3 به صورت

$$\delta_3 = \frac{(X_2 - X_1)^T}{\lambda_1^*} \times \frac{A \nabla f_3}{S_1^T S_1} \quad (32-2)$$

به دست می‌آید. رابطه (۳۲-۲) با استفاده از رابطه (۲۷-۲) به صورت

$$\delta_3 \frac{1}{\lambda_1^*} \times \frac{(\nabla f_2 - \nabla f_1)^T \nabla f_3}{S_1^T A S_1} \quad (33-2)$$

حاصل می‌شود. از روابط (۱۸-۲) و (۲۲-۲) داریم:

$$S_1 = -\nabla f_1$$

$$S_2 - \beta_2 S_1 = -\nabla f_2$$

در نتیجه

$$\nabla f_2 - \nabla f_1 = -S_2 + S_1(1 + \beta_2) \quad (34-2)$$

از رابطه (۳۳-۲) داریم:

$$\delta_3 = \frac{1}{\lambda_3^*} \times \frac{\{-S_2 + (1 + \beta_2)S_1\}^T \nabla f_3}{S_1^T A S_1} \quad (35-2)$$

با توجه به رابطه (۱۷-۲) مشاهده می‌شود که مقدار عبارت فوق برابر با صفر می‌شود. در نتیجه رابطه (۲۹-۲) به صورت

$$S_3 = -\nabla f_3 + \beta_3 S_2 \quad (36-2)$$

به دست می‌آید. لازم به ذکر است که مقدار β_3 با مزدوج قرار دادن S_3 با S_2 نتیجه می‌شود.

در حالت کلی برای یافتن β_i , ($i = 2, 3, \dots$) به صورت زیر عمل می‌کنیم.

با تعمیم رابطه (۳۶-۲)، S_i ، جهت جستجو در i -امین گام را به عنوان یک ترکیب خطی از $-\nabla f_i$ و S_{i-1} بیان می‌کنیم.

به عبارت دیگر

$$S_i = -\nabla f_i + \beta_i S_{i-1} \quad (37-2)$$

که مقدار β_i از مزدوج قرار دادن S_i با S_{i-1} به صورت

$$\beta_i = \frac{\nabla f_i^T \nabla f_i}{\nabla f_{i-1}^T \nabla f_{i-1}}$$

حاصل می‌شود.

تا اینجا S_i و S_{i-1} را با یکدیگر مزدوج در نظر گرفتیم.

حال نشان می‌دهیم که S_i به دست آمده از رابطه (۲-۳۷) به این شرط که S_1, S_2, \dots, S_{i-1} مزدوج باشند، خود به خود با همه S_k ها مزدوج است. رابطه را در نظر بگیرید:

$$\begin{aligned} S_k^T A S_i &= S_k^T A (-\nabla f_i + \beta_i S_{i-1}) \\ &= -S_k^T A \nabla f_i + \beta_i S_k^T A S_{i-1} \quad , \quad k = 1, 2, \dots, i-2 \end{aligned} \quad (۲-۳۸)$$

از آنجایی که $S_k^T A S_{i-1} = 0$ و $k = 1, 2, \dots, i-2$ داریم:

$$S_k^T A S_i = -S_k^T A \nabla f_i \quad , \quad k = 1, 2, \dots, i-2 \quad (۲-۳۹)$$

روابط مشابه با (۲-۲۵) و (۲-۲۷) را می‌توان به صورت

$$S_k = \frac{(X_{k+1} - X_k)}{\lambda_k^*} \quad (۲-۴۰)$$

$$\nabla f_{k+1} - \nabla f_k = A(X_{k+1} - X_k) \quad (۲-۴۱)$$

به دست آورد. با توجه روابط

$$\nabla f_k^T \nabla f_{i+1} = 0 \quad , \quad k = 1, 2, \dots, i \quad (۲-۴۲)$$

$$S_k = -\nabla f_k + \beta_k S_{k-1}$$

رابطه (۲-۱۷) به فرم

$$S_k^T \nabla f_{i+1} = -\nabla f_k^T \nabla f_{i+1} + \beta_k S_{k-1}^T \nabla f_{i+1} = 0 \quad , \quad k = 1, 2, \dots, i$$

نتیجه می‌شود. با توجه به رابطه داریم: (۲-۱۷)

$$S_{k-1}^T \nabla f_{i+1} = 0$$

- 1: $g_0 := -\nabla f(x_0)$
- 2: $h_0 := g_0$
- 3: **for** $k = 1$ to K **do**
- 4: $\alpha_k := \min_{\alpha} f(x_{k-1} + \alpha h_{k-1})$
- 5: $x_k := x_{k-1} + \alpha_k h_{k-1}$
- 6: $g_k := -\nabla f(x_k)$
- 7: $h_k := g_k + \frac{g_k \cdot g_k}{g_{k-1} \cdot g_{k-1}} h_{k-1}$
- 8: **end for**
- 9: $x := x_k$

بنابراین

$$\nabla f_k^T \nabla f_{i+1} = 0, \quad k = 1, 2, \dots, i$$

در نتیجه رابطه (۲-۳۹) به صورت زیر

$$S_k A S_i = \frac{-1}{\lambda_k^*} (\nabla f_{k+1} = \nabla f_k)^T \nabla f_i = 0, \quad k = 1, 2, \dots, i-2 \quad (۲-۴۳)$$

به دست می‌آید.

الگوریتم پایه مینیمم‌سازی گرادیان مزدوج برای تابع $f(x)$ در الگوریتم ۲-۲ آورده شده است.

۲-۸ به‌کارگیری روش گرادیان مزدوج برای افرازبندی طیفی گراف

روش گرادیان مزدوج، یکی از متداول‌ترین روش‌ها برای تعیین کوچک‌ترین (بزرگ‌ترین) مقدار ویژه و بردار ویژه متقارن است. ایده اصلی این روش مینیمم‌سازی خارج قسمت رایلی به وسیله روش گرادیان مزدوج است. روش گرادیان مزدوج یک روش کلی شناخته شده برای مینیمم‌سازی توابع است [۱۸].

در این روش خارج قسمت رایلی که به صورت

$$F(x) = \frac{x \cdot L \cdot x}{x \cdot x} \quad (۲-۴۴)$$

تعریف می‌شود باید مینیمم شود. همچنین گرادیان خارج قسمت رایلی به صورت

$$\nabla F(x) = \frac{2}{x \cdot x} (L \cdot x - F(x)x). \quad (۲-۴۵)$$

۸-۱-۲ روش مینیمم سازی گرادیان مزدوج برای افراز بندی طیفی گراف

کوچکترین مقدار ویژه یک ماتریس متقارن را می توان با مینیمم سازی خارج قسمت رایلی نسبت به تمام x ها به دست آورد. همچنین دومین-کوچکترین مقدار ویژه را می توان با مینیمم سازی خارج قسمت رایلی نسبت به تمام x هایی که بر بردار ویژه متناظر با کوچکترین مقدار ویژه عمود است به دست آورد.

از آن جایی که این بردار ویژه برابر با e است، بنابراین مینیمم سازی باید تمام x هایی را برگرداند که $x \cdot e = 0$ است. اگر تخمین اولیه x برای بردار ویژه بر e عمود باشد، در این صورت تمام تخمین های x_k هایی که با روش مینیمم سازی گرادیان مزدوج تولید می شوند بر e عمودند. بنابراین روش مینیمم سازی گرادیان مزدوج، برای تعیین کوچکترین مقدار ویژه را می توان برای تعیین دومین-کوچکترین مقدار ویژه نیز استفاده کرد، که به سادگی با استفاده از تخمین اولیه ای که بر e عمود است، انجام می شود. می توان از بردار $x_i = i - (|V| + 1) / 2$ ، که $|V|$ برابر با تعداد راس های گراف متناظر با ماتریس لاپلاسیان L است، به عنوان این تخمین اولیه استفاده کرد [۲۰]. به وضوح شرط $x \cdot e = 0$ برای این تخمین اولیه برقرار است.

در اینجا تغییرات اصلی الگوریتم گرادیان مزدوج پایه بخش قبل، برای مینیمم سازی تابع خارج قسمت رایلی ارائه می شود. این تغییرات مربوط به مینیمم سازی خطی در سطر (۴) و محاسبه گرادیان در سطر (۶) الگوریتم ۲-۲ می باشند. با توجه به فرم خاصی از خارج قسمت رایلی، مینیمم سازی خطی به صورت

$$\min_{\alpha} F(x + \alpha h) \equiv \min_{\alpha} \frac{x \cdot L \cdot x + 2(x \cdot L \cdot h)\alpha + (h \cdot L \cdot h)\alpha^2}{x \cdot x + 2(x \cdot h)\alpha + (h \cdot h)\alpha^2}. \quad (46-2)$$

نتیجه می شود. α از یک معادله درجه دوم بدست می آید که این α خارج قسمت رایلی را مینیمم می کند. عبارت $x \cdot L \cdot x$ برابر با $F(x) \cdot x \cdot x$ است. بنابراین تنها باید یک ضرب ماتریس-بردار $L \cdot h$ محاسبه شود تا مینیمم سازی خطی بدست آید. ارزیابی مستقیم گرادیان از (۲-۴۵) به یک ضرب اضافی ماتریس-بردار نیاز دارد. راه حل مؤثری برای محاسبه گرادیان با در نظر گرفتن رابطه به روزرسانی سطر (۵) الگوریتم ۲-۲ بدست می آید که منجر به رابطه به روزرسانی زیر برای بردار g می شود [۱۹]:

$$g_k := \frac{2}{x_k \cdot x_k} \left(F_k x_k - F_{k-1} x_{k-1} + \frac{1}{4} (x_{k-1} \cdot x_{k-1}) g_{k-1} - \alpha_k L \cdot h_{k-1} \right). \quad (47-2)$$

اگر $x_k \cdot e = 0$ ، آنگاه از (۱۱-۱۳) و $L \cdot e = 0$ نتیجه می شود که $g_k \cdot e = 0$ و $h_k \cdot e = 0$. این بدان معناست که با استفاده از سطرهای (۵) و (۷) الگوریتم ۲-۲ و معادله (۲-۴۷)، $x_k \cdot e = 0$ ، $g_k \cdot e = 0$ و $h_k \cdot e = 0$ در نتیجه اثبات می شود که اگر تخمین اولیه برای بردار ویژه متناظر با دومین-کوچکترین مقدار ویژه بر e عمود باشد سپس تمام برآوردهای دیگر که به وسیله روش مینیمم سازی گرادیان مزدوج بدست می آیند نیز بر e عمودند.

الگوریتم ۲-۳ روش گرادیان مزدوج برای افزایش بندی طیفی گراف [۱۹]

```

1:  $x_0 := x_0 - \left(\frac{e \cdot x_0}{e \cdot e}\right) e$ 
2:  $\rho_0 := x_0 \cdot x_0$ 
3:  $y_0 := L \cdot x_0$ 
4:  $F_0 := \frac{x_0 \cdot y_0}{\rho_0}$ 
5:  $g_0 := \frac{2}{\rho_0} (F_0 x_0 - y_0)$ 
6:  $h_0 := g_0$ 
7:  $\beta_0 := g_0 \cdot g_0$ 
8: for  $k = 1$  to  $K$  do
9:    $y_k := L \cdot h_{k-1}$ 
10:   $p_k := \rho_{k-1}$ 
11:   $q_k := 2x_{k-1} \cdot h_{k-1}$ 
12:   $r_k := h_{k-1} \cdot h_{k-1}$ 
13:   $s_k := \rho_{k-1} F_{k-1}$ 
14:   $t_k := 2x_{k-1} \cdot y_k$ 
15:   $u_k := h_{k-1} \cdot y_k$ 
16:   $\alpha_k := \frac{-(u_k p_k - s_k r_k) + \sqrt{(u_k p_k - s_k r_k)^2 - (u_k q_k - t_k r_k)(t_k p_k - s_k q_k)}}{u_k q_k - t_k r_k}$ 
17:   $F_k := \frac{s_k + t_k \alpha_k + u_k \alpha_k^2}{p_k + q_k \alpha_k + r_k \alpha_k^2}$ 
18:   $x_k := x_{k-1} + \alpha_k h_{k-1}$ 
19:   $\rho_k := x_k \cdot x_k$ 
20:   $g_k := \frac{2}{\rho_k} (F_k x_k - F_{k-1} x_{k-1} + \frac{1}{2} \rho_{k-1} g_{k-1} - \alpha_k y_k)$ 
21:   $\beta_k := g_k \cdot g_k$ 
22:   $h_k := g_k + \frac{\beta_k}{\beta_{k-1}} h_{k-1}$ 
23: end for
24:  $x := x_k$ 

```

در الگوریتم ۲-۳ پیاده سازی روش گرادیان مزدوج برای افزایش بندی طیفی گراف ارائه شده است. در هر تکرار باید سه بردار برورسانی، یک ضرب ماتریکس-بردار و ۶ ضرب بردار-بردار انجام شوند. فضای حافظه مورد نیاز به دلیل ذخیره سازی تنها چهار بردار، کم است.

۲-۹ روش نیستروم برای تخمین مقدار ویژه و بردار ویژه

در روش نیستروم به جای استفاده از تمام n داده، تعدادی از داده ها به عنوان نمونه انتخاب می شوند و داده ها به دو بخش تقسیم می شوند، یک بخش شامل m نقطه نمونه تصادفی بوده و بخش دیگر شامل $n - m$ نقطه باقی مانده است. بنابراین

ماتریس مجاورت W را می‌توان به صورت زیر نوشت:

$$W = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \quad (48-2)$$

جایی که $A \in \mathbb{R}^{m \times m}$ ماتریس مجاورت نقاط نمونه و $A = U\Lambda U^T$ ؛ $B \in \mathbb{R}^{m \times (n-m)}$ ماتریس مجاورت نقاط نمونه و نقاط باقی‌مانده و $C \in \mathbb{R}^{(n-m) \times (n-m)}$ ماتریس مجاورت نقاط باقی‌مانده باشد. با استفاده از روش نیستروم \bar{U} به عنوان تخمینی برای بردار ویژه ماتریس W به صورت زیر به دست می‌آید [21]:

$$\bar{U} = \begin{bmatrix} U \\ B^T U \Lambda^{-1} \end{bmatrix}$$

به صورت متناظر

$$\begin{aligned} \hat{W} &= \bar{U} \Lambda \bar{U}^T \\ &= \begin{bmatrix} U \\ B^T U \Lambda^{-1} \end{bmatrix} \Lambda [U^T \Lambda^{-1} U^T B] \\ &= \begin{bmatrix} U \Lambda U^T & B \\ B^T & B^T \Lambda^{-1} B \end{bmatrix} \\ &= \begin{bmatrix} A & B \\ B^T & B^T \Lambda^{-1} B \end{bmatrix} \end{aligned}$$

در روش نیستروم از $B^T \Lambda^{-1} B$ به عنوان تقریبی برای C استفاده می‌شود. اگر $m \ll n$ ، تعداد نقاط باقی‌مانده پس از نمونه‌گیری بسیار بزرگ است. روش تقریبی نیستروم از محاسبه ماتریس مجاورت نقاط باقی‌مانده اجتناب می‌کند و بنابراین به میزان قابل توجهی پیچیدگی فضا و زمان الگوریتم را کاهش می‌دهد. گرچه نمی‌توان مستقیماً از بردار ویژه \bar{U} استفاده کرد، چون لزوماً با ترانزاده‌اش متعامد نیست. قضیه ۱-۹-۲ صورت بردار ویژه متعامد \hat{W} را ارائه می‌دهد.

قضیه ۱-۹-۲. اگر A معین مثبت باشد، $A^{1/2}$ ریشه متقارن و معین مثبت A باشد، $Q = A + A^{-1/2} B B^T A^{-1/2}$ را تعریف کرده و تجزیه آن $Q = U_Q \Lambda_Q U_Q^T$ ، آنگاه بردار ویژه متعامد \hat{W} به صورت زیر است:

$$V = \begin{bmatrix} A \\ B^T \end{bmatrix} A^{-1/2} U_Q \Lambda_Q^{-1/2} \quad (49-2)$$

برهان. ابتدا ثابت می‌کنیم V بردار ویژه ماتریس \hat{W} است.

$$\begin{aligned}\hat{W} &= \begin{bmatrix} A & B \\ B^T & B^T A^{-1} B \end{bmatrix} \\ &= \begin{bmatrix} A \\ B^T \end{bmatrix} A^{-1} \begin{bmatrix} A & B \end{bmatrix} \\ &= \left\{ \begin{bmatrix} A \\ B^T \end{bmatrix} A^{-1/2} U_Q \Lambda_Q^{-1/2} \right\} \Lambda_Q \times \left\{ \Lambda_Q^{-1/2} U_Q^T A^{-1/2} \begin{bmatrix} A & B \end{bmatrix} \right\} \\ &= V \Lambda_Q V^T\end{aligned}$$

حال اثبات می‌کنیم V و V^T متعامدند

$$\begin{aligned}I &= V^T V \\ &= \left\{ \Lambda_Q^{-1/2} U_Q^T A^{-1/2} \begin{bmatrix} A & B \end{bmatrix} \right\} \left\{ \begin{bmatrix} A \\ B^T \end{bmatrix} A^{-1/2} U_Q \Lambda_Q^{-1/2} \right\}\end{aligned}$$

با ضرب دو بخش سمت راست معادله در $U_Q \Lambda_Q^{1/2}$ و $\Lambda_Q^{1/2} U_Q^T$ مقدار Q را به دست می‌آوریم.

$$\begin{aligned}Q &= U_Q \Lambda_Q U_Q^T \\ &= A^{-1/2} \begin{bmatrix} A & B \\ B^T & B^T A^{-1} B \end{bmatrix} A^{-1/2} \\ &= A + A^{-1/2} B B^T A^{-1/2}\end{aligned}$$

□

برای استفاده از روش نیستروم در خوشه‌بندی طیفی، باید ماتریس مجاورت یا ماتریس لاپلاسیان را نرمال‌سازی^۱ کنیم، بدین ترتیب $\hat{D}^{-1/2} \hat{W} \hat{D}^{-1/2}$ ، جایی که \hat{D} ماتریس قطری است که \hat{D}_{ii} برابر با جمع i امین سطر \hat{W} باشد. فولکس^۲

^۱Normalization

^۲Fowlkes

و همکاران یک روش محاسباتی ساده را برای محاسبه درجه گره ارائه دادند [۲۲]:

$$\begin{aligned} \hat{d} = \hat{W}^{-1} &= \begin{bmatrix} A^{-1} \mathbf{1}_m + B^{-1} \mathbf{1}_{n-m} \\ B^T \mathbf{1}_m + B^T A^{-1} B \mathbf{1}_{n-m} \end{bmatrix} \\ &= \begin{bmatrix} a_r + b_r \\ b_c + B^T A^{-1} b_r \end{bmatrix} \end{aligned} \quad (50-2)$$

جایی که $a_r, b_r \in \mathbb{R}^m$ نشان‌دهنده جمع سطری A و B ، $b_c \in \mathbb{R}^{n-m}$ نشان‌دهنده جمع ستونی B ، $\mathbf{1}$ بردار ستونی با تمام مقادیر ۱ باشند. ماتریس‌های A و B را می‌توان با \hat{d} نرمال‌سازی کرد:

$$A_{ij} \leftarrow \frac{A_{ij}}{\sqrt{\hat{d}_i \hat{d}_j}}, \quad i, j = 1, \dots, m \quad (51-2)$$

$$B_{ij} \leftarrow \frac{B_{ij}}{\sqrt{\hat{d}_i \hat{d}_{j+m}}}, \quad i, j = 1, \dots, m, \quad j = 1, \dots, n - m \quad (52-2)$$

۱۰-۲ روش نمونه‌گیری افزایشی بر پایه احتمال

خوشه‌بندی طیفی، داده‌ها را از فضای اصلی به فضای-ویژه با بعد کمتر تصویر می‌کند تا داده‌هایی با ساختار پیچیده بصورت خطی جداپذیر شوند. در این روش مقادیر ویژه و بردارهای ویژه ماتریس لاپلاسیان مورد نیاز است که هزینه محاسباتی و فضای مورد نیاز برای محاسبه این مقادیر و نگهداری ماتریس مجاورت برای استفاده‌های امروزی بالا است. یک راه‌حل مورد استفاده برای کاهش این هزینه‌ها استفاده از روش نیستروم برای تقریب زدن مقادیر ویژه و بردارهای ویژه است. در این روش از داده‌ها تعدادی بعنوان نمونه انتخاب می‌شوند و از این نمونه‌ها برای تقریب زدن استفاده می‌شود. همانگونه که انتظار می‌رود چگونگی انتخاب نمونه‌ها تأثیر زیادی در کیفیت تقریب دارد. اولین روشی که به ذهن می‌رسد، روش تصادفی است که کم‌هزینه‌ترین روش است اما این روش پایداری و ثبات لازم را نداشته و همچنین نمونه‌گیری یکسان از یک مجموعه داده بزرگ سخت بوده و احتمال از دست دادن داده‌های مؤثر وجود دارد. اگر نمونه‌ها نمایندگان خوبی از مجموعه داده نباشند تأثیر خوبی بر خوشه‌بندی نمی‌گذارند.

در روش مبتنی بر احتمال یک تابع احتمالی معرفی می‌شود که انتخاب نمونه‌ها را هوشمندانه‌تر انجام می‌دهد. ایده اصلی خوشه‌بندی طیفی محاسبه k بردار ویژه اولیه ماتریس لاپلاسیان است تا یک زیرفضا با بعد کمتر ایجاد کند که در آن ساختار داده‌ها بهینه شده و مرز میان خوشه‌ها مشخص‌تر شده باشد. بدلیل هزینه محاسباتی بالا تجزیه-ویژه از روش نیستروم برای تقریب بردارهای ویژه ماتریس لاپلاسیان استفاده می‌کنیم.

هر ماتریس A تعداد k/ε سطر دارد که فضای تولید شده توسط این سطرها می تواند یک تقریب رتبه- k درست کند که خطای افزوده آن کمتر از $\varepsilon \|A\|_F^2$ باشد [۷]. زیر مجموعه سطرهای انتخاب شده به عنوان نمونه های مستقل از توزیع احتمالی که با نرم سطرهای A تصمیم گیری شده است، بدست می آید.

قضیه ۱-۱۰-۲. ماتریس $A \in \mathbb{R}^{m \times n}$ را داریم و می خواهیم s سطر از A را انتخاب کنیم تا ماتریس S را بسازیم، و احتمال انتخاب سطر i ام برابر است با:

$$P_i = \frac{\|A^{(i)}\|_2^2}{\|A\|_F^2}$$

که $A^{(i)}$ سطر i ام ماتریس A است. اگر $s \geq k/\varepsilon$ ، ماتریس \tilde{A}_k (بیشترین رتبه k) با سطرهایش که مشمول در فضای گسترش یافته (S) است نامساوی زیر را ارضا می کند

$$E_S \left(\|A - \tilde{A}_k\|_F^2 \right) \leq \|A - A_k\|_F^2 + \varepsilon \|A\|_F^2$$

قضیه ۱-۱۰-۲ می تواند به یک الگوریتم نمونه گیری تبدیل شود [۲۳]. ابتدا با پیمایش سطرهای A توزیع احتمالی نمونه ها را بدست می آید، سپس نمونه هایی که بیشترین احتمال را دارند انتخاب می شوند. پیچیدگی این الگوریتم برابر

$$O(\min\{m, n\}k^2/\varepsilon^2)$$

است. خطای تخمین با چندین بار پیمایش کردن داده ها کاهش می یابد. به طور مثال، اگر تمام داده ها در امتداد یک زیر فضای یک بعدی از \mathbb{R}^n باشند بجز یکی از داده ها، بهترین زیر فضای مرتبه ۲ باید خطای صفر داشته باشد. احتمال این که در دور اول نقطه معجزا انتخاب نشود وجود دارد. دور دوم بهتر می شود. در دور اول مجموعه ای از نمونه ها را با استفاده از توزیع مربع نرم بدست می آوریم؛ سپس احتمال نمونه گیری داده های باقی مانده را محاسبه می کنیم، احتمال انتخاب (نمونه گیری) متناسب با مربع فاصله بین نقاط و فضای اسپن شده با نمونه های موجود است. گروه دیگری از نمونه ها را با احتمال گفته شده انتخاب می کنیم. این پروسه نمونه گیری افزایشی براساس احتمال نامیده می شود. اگر نقطه دور افتاده (جداشده) در نمونه گیری اول انتخاب نشود، احتمال انتخاب آن در مرحله دوم نمونه گیری بالا است. اکنون فضای تولید شده می تواند یک تقریب خوب رتبه-۲ درست کند. با افزایش تعداد پیمایش ها خطا نیز به صورت نمایی کاهش می یابد. پیمایش چندین باره داده ها برای مسئله تقریب رتبه-پایین بسیار مفید است. توصیف ریاضی نمونه گیری افزایشی در قضیه زیر آمده است.

قضیه ۲-۱۰-۲. فرض کنید $S = S_1 \cup \dots \cup S_t$ نشان دهنده مجموعه سطرهای تصادفی انتخاب شده ماتریس $m \times n$ A ، باشد. برای $j = 1, \dots, t$ ، هر S_j یک مجموعه نمونه از s سطر A است که به طور مستقل از توزیع زیر انتخاب

شده‌اند، احتمال انتخاب سطر i برابر است با

$$P_i^{(j)} = \frac{\|E_j^{(i)}\|_F^2}{\|E_j\|_F^2} \quad (53-2)$$

که $E_1 = A$ ، $E_j = A - \pi_{S_1 \cup \dots \cup S_{j-1}}(A)$ ، $\pi_S(A)$ نشان دهنده تصویر A بر فضای تولید شده S است. برای ماتریس \tilde{A}_k (بیشترین رتبه k است) با سطرهایش در فضای تولید شده S مشمول است نامساوی زیر را ارضا می‌کند

$$E_S \left(\|A - \tilde{A}_k\|_F^2 \right) \leq \frac{1}{1-\varepsilon} \|A - A_k\|_F^2 + \varepsilon^T \|A\|_F^2$$

در قضیه ۲-۱۰-۲ می‌توان دید که توزیع نمونه‌ها t بار تغییر می‌کند اما خود ماتریس تغییر نمی‌کند، وقتی A اسپارس باشد، حفظ اسپارس بودنش بسیار مهم است.

۱۱-۲ الگوریتم افرازبندی طیفی نیستروم مبتنی بر نمونه‌گیری احتمالی افزایشی

در بخش ۲-۱۰، نمونه‌گیری افزایشی مبتنی بر احتمال به صورت تئوری بررسی شد. بیشتر روش‌های نمونه‌گیری، نمونه‌ها را از یک توزیع ثابت انتخاب می‌کنند. نمونه‌گیری افزایشی احتمال انتخاب همه داده‌ها را بعد از انتخاب هر گروه از نمونه به‌روزرسانی خواهد کرد به طوری که نمونه‌های به دست آمده می‌توانند تقریبی از زیرفضای بهینه مرتبه- k که توسط $\text{span}\{v^{(1)}, \dots, v^{(k)}\}$ بوجود می‌آید را تولید کنند. با این روش نمونه‌گیری افزایشی می‌توان چندین سطر از ماتریس مجاورت W را انتخاب کرد و ماتریس تقریبی مجاورت \hat{W} را ساخت. نمونه‌گیری افزایشی از توزیع اولیه تمام سطرهای W آغاز می‌شود: ابتدا $s < m$ سطر را با حداکثر احتمال برای تشکیل زیر ماتریس R' ، انتخاب کنید. سپس احتمال انتخاب سطرهای باقی‌مانده را با استفاده از R' به‌روز کنید. s سطر دیگر از W را با توجه به احتمال نمونه‌گیری جدید انتخاب کنید و آنها را در R' قرار دهید؛ این فرایند را تا زمانی که m سطر انتخاب شدند، تکرار کنید. جزئیات طرح نمونه‌گیری افزایشی در الگوریتم ۲-۴ نشان داده شده است.

پس از به دست آوردن m اندیس از سطرهای نمونه از طریق الگوریتم ۲-۴، می‌توان با استفاده از روش نیستروم بردارهای ویژه ماتریس لاپلاسین نرمال شده را تقریب زد و زمان محاسباتی افرازبندی طیفی را کاهش داد. در الگوریتم ۲-۵ تمام روند الگوریتمیک این روش آورده شده است.

الگوریتم ۲-۴ الگوریتم نمونه‌گیری افزایشی بر پایه احتمال (Incremental Sampling) [۲۸]

Require: the similarity matrix $W \in \mathbb{R}^{n \times n}$ of data point, the total number of sampling rows m , the number of selected rows s in each iteration

Ensure: the set S , m indexes of sampled rows in W

- 1: Initialize the index set $S = \emptyset$ of samples, set the number of iterations $t = m/s$.
 - 2: **for** $i \in [1 \dots t]$ **do**
 - 3: $P_i \leftarrow \text{Update Probability}(S)$
 - 4: $S_i \leftarrow$ the set of s selected indexes according to P_i
 - 5: $S \leftarrow S \cup S_i$
 - 6: **end for**
-

الگوریتم ۲-۵ بروزرسانی احتمال (Update Probability) [۲۸]

Require: Sample indexes set S

Ensure: The probability set P

- 1: From set R' with the selected rows of W corresponding to the indexes in S .
 - 2: Calculate the error between W and its orthogonal projection on R' :
$$E = W - U_{R'} W U_{R'}^T.$$
 - 3: **for** $j \in [1 \dots n]$ **do**
 - 4: if $j \in S$, then let $P_j = 0$;
 - 5: Otherwisr, let $P_j = \|E_j\|_2^2$.
 - 6: **end for**
 - 7: Update the sampling probabilities of all rows $P \leftarrow \frac{P}{\|P\|_2}$.
-

الگوریتم ۲-۶ افزایش برابری طیفی نیستروم مبتنی بر نمونه‌گیری احتمالی افزایشی [۲۱].

Require: the data set $X = \{x_i | i = 1, \dots, n\}$, the number of samples m , the number of clusters k .

Ensure: k grouped clusters.

- 1: Calculate the pairwise similarity between data points in X to construct similarity matrix $W \in \mathbb{R}^{n \times n}$, where each element w_{ij} can be expressed by Gaussian kernel, namely $w_{ij} = \exp(-|x_i - x_j|^2 / 2\sigma^2)$.
 - 2: Select m row samples of W by Algorithm 4-2 to compose the similarity matrix $A \in \mathbb{R}^{m \times m}$ between sampled points and the remaining points, and the similarity matrix $B \in \mathbb{R}^{m \times (n-m)}$ between the sampled points and the remaining points.
 - 3: Compute the degree \hat{d} of nodes according to formula (50-2) based on matrix A and B , then normalize matrix A and B respectively with formula (51-2) and (52-2).
 - 4: Calculate matrix Q with the normalized A and B : $Q = A + A^{-1/2} B B^T A^{-1/2}$.
 - 5: Get matrix U_Q and Λ_Q by decomposing matrix Q : $Q = U_Q \Lambda_Q U_Q^T$.
 - 6: Substitute U_Q , Λ_Q into formula (49-2) to calculate the orthogonal eigenvector V of \hat{W} .
 - 7: Select the eigenvectors corresponding to the first k largest eigenvalues v_1, \dots, v_k of \hat{W} from V to form matrix V_k : $V_k = [v_1 : \dots : v_k] \in \mathbb{R}^{n \times k}$.
 - 8: Normalize each row of matrix V_k to a unit vector and obtain matrix Y with each element $y_{ij} = v_{ij} / \sqrt{\sum_{j=1}^k v_{ij}^2}$.
 - 9: Matrix Y forms a low-dimensional embedding space $\mathbb{R}^{n \times k}$, in which the i -th row is corresponding to point x_i . We may use k-means algorithm to classify the row vectors of Y into k clusters, then restore the clustering results data then restore the clustering results of the original data points.
-

فصل ۳

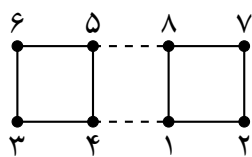
نتایج آزمایشات و روش‌های پیشنهادی

۱-۳ استفاده از الگوریتم ژنتیک در گرادیان مزدوج

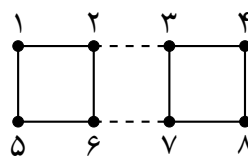
در صورت نامناسب بودن نقطه شروع اولیه در روش گرادیان مزدوج، این شیوه ممکن است در کمینه محلی گرفتار شود. این مشکل در مسئله افراز گراف نیز وجود دارد. شکل ۳-۱ یک گراف که در آزمایشات انجام شده، روش گرادیان مزدوج، با این مشکل مواجه شده است را نشان می‌دهد. شکل ۳-۱(آ) یک گراف با ۸ رأس را نشان می‌دهد که هر دو روش خوشه‌بندی طیفی و گرادیان مزدوج پاسخ درست را می‌دهند. کمینه تعداد یال‌های برشی که با قطع آنها گراف به دو دسته با تعداد رئوس نسبتاً برابر تقسیم می‌شود، ۲ یال است که در شکل مذکور با خط چین نشان داده شده است. شکل ۳-۱(ب) همین گراف را با ترتیب دیگری از رئوس نشان می‌دهد. خوشه‌بندی طیفی، نتیجه درست را می‌دهد که در آن یال‌های مورد برش همان ۲ یال نشان داده شده با خط چین هستند، اما روش گرادیان مزدوج خروجی نمایش داده شده در شکل ۳-۱(ج) را تولید می‌کند که اشتباه است. ترتیب رؤس در این گراف به نحوی هستند که نقطه شروع الگوریتم گرادیان مزدوج، نقطه مناسبی نباشد. در این بخش روشی برای پیدا کردن نقطه مناسب این الگوریتم با روش الگوریتم ژنتیک ارائه می‌شود.

به دلیل همگرایی الگوریتم گرادیان مزدوج به کمینه محلی، انتخاب نقطه اولیه در رسیدن به بهینه سراسری حائز اهمیت است. بردارهای ویژه ماتریس لاپلاسیان، که جواب درست مسئله هم جزو آن‌هاست، با یکدیگر متعامند [۱۹]. در نتیجه ضرب داخلی آن‌ها برابر صفر می‌باشد. با انتخاب نقطه اولیه x به گونه‌ای که $(\forall i = 1, 2, \dots, |V| = n) x_i = i - (|V| + 1)/2$ به دلیل مزدوج بودن جهات در روش گرادیان مزدوج تمام نقاط بدست آمده بعدی نیز با بردار e متعامند [۱۲].

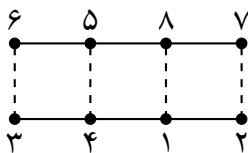
اگر درایه‌های x جابجا شوند، همچنان x با بردار e متعامد باقی می‌ماند. می‌توان با جابجایی این درایه‌ها نقطه اولیه جدیدی تولید کرد به این امید که الگوریتم در کمینه محلی گرفتار نشود. روش پیشنهادی در این مقاله، استفاده از الگوریتم ژنتیک برای تولید جایگشت‌های متفاوت از بردار شامل رئوس گراف است. هر جایگشت به عنوان یک کروموزوم، یک جواب



(ب) خروجی مورد انتظار برای رأس‌های با شماره‌های ۶، ۵، ۸، ۷، ۳، ۴، ۱، ۲.



(آ) خروجی درست با رأس‌های با شماره‌های ۱ تا ۸.



(ج) خروجی نادرست روش گرادیان مزدوج

شکل ۳-۱: مثال‌های از عملکرد درست و نادرست الگوریتم گرادیان مزدوج در افراز گراف. خط‌چین‌ها معرف لبه‌های مورد برش در گراف هستند.

اولیه متفاوت را تولید می‌کند. ترکیب جواب‌های هر نسل و تولید نسل بعد بر اساس میزان بهینه بودن جواب حاصل از الگوریتم گرادیان مزدوج، مطابق با شیوه کلی کار الگوریتم ژنتیک، نتایج مناسبی را خواهد داد. برای ترکیب کرموزوم‌ها از روش بازترکیب مرتبه ۱ و برای جهش از عملگر تعویض دوتایی استفاده شده است. هر دو عملگر بکار برده شده مناسب مسائل جایگشتی هستند. تعداد اعضای جمعیت ۲۰ و حداکثر تعداد نسل‌ها، ۱۰ در نظر گرفته شده است.

به منظور بررسی کارایی شیوه پیشنهادی در بهبود روش گرادیان مزدوج از طریق جستجوی نقطه شروع اولیه مناسب، چندین گراف تصادفی با ابعاد مختلف تولید شده و در آزمایشات بکار گرفته شده‌اند. نتایج آزمایشات در جدول‌های ۳-۱ تا ۳-۳ آمده است. شماره ردیف در هر جدول بیانگر شماره مثال مورد استفاده است. گراف‌های مورد بررسی، گراف‌های تصادفی هستند که به ترتیب دارای ۱۵۰، ۲۰۰، ۲۵۰، ۳۰۰ و ۳۵۰ رأس می‌باشند. تابع هدف مورد استفاده در الگوریتم ژنتیک، تابع (۷-۲) بوده است که نتایج مقایسه‌ای بر مبنای این تابع هدف در جدول ۳-۱ آمده است. برای هر اندازه‌ی گراف، ۲۰ گراف تصادفی تولید شده و میانگین نتایج در جداول آمده است. به این ترتیب به عنوان نمونه، سطر اول جدول فوق‌الذکر، نتایج اجرا بر روی ۲۰ گراف تصادفی با ۱۵۰ رأس می‌باشد. سطر آخر هر جدول میانگین هر ستون را نشان می‌دهد. همان‌گونه که در شکل ۳-۱ دیده می‌شود، الگوریتم ژنتیک، توانسته است در تمام مثال‌ها نتایج روش گرادیان مزدوج را بهبود بخشد.

با خوشه‌بندی بردار x حاصل از گرادیان مزدوج می‌توان بردار جدید شامل $+1$ و -1 را بدست آورد که با قرار دادن این بردار در رابطه $\frac{1}{2}x^t Lx$ از لم ۲-۱-۱ تعداد یال‌های برشی قابل محاسبه است. گراف‌های تولید شده در آزمایشات انجام شده، گراف‌هایی بدون وزن و بدون جهت بوده‌اند. لذا در ماتریس مجاورتی گراف، وجود لبه با یک و عدم وجود لبه با صفر نشان داده شده است. از این‌رو، مجموع تعداد لبه‌های برشی بین دو خوشه، محکی برای میزان بهینه بودن خوشه‌بندی انجام شده می‌باشد. جدول ۳-۲ تعداد لبه‌های برشی بین دو خوشه در دو روش مورد بررسی را نشان می‌دهد. همان‌گونه که انتظار

روش		شماره
GACG	CG	
۳/۱۷۳	۰/۰۸۰	۱
۲/۷۶۰	۰/۰۰۷	۲
۳/۲۲۲	۰/۰۰۹	۳
۳/۷۸۸	۰/۰۱۲	۴
۴/۳۹۱	۰/۰۱۳	۵
۳/۴۶۷	۰/۰۲۴	میانگین

روش		شماره
GACG	CG	
۳۱۳/۰۰۰	۳۳۴/۲۰۰	۱
۴۵۹/۹۰۰	۴۹۰/۶۵۰	۲
۶۷۹/۵۰۰	۷۲۶/۴۰۰	۳
۹۵۷/۵۵۰	۹۹۹/۸۰۰	۴
۱۲۸۸/۶۰۰	۱۳۳۶/۹۵۰	۵
۷۳۹/۷۱۰	۷۷۷/۶۰۰	میانگین

روش		شماره
GACG	CG	
۷/۷۶۶	۸/۹۹۱	۱
۸/۷۸۵	۹/۸۶۳	۲
۱۰/۴۱۴	۱۱/۷۲۶	۳
۱۲/۱۴۳	۱۳/۲۴۸	۴
۱۴/۲۴۹	۱۵/۲۰۰	۵
۱۰/۶۷۲	۱۱/۸۰۵	میانگین

جدول ۳-۳: زمان اجرا الگوریتم گرادیان مزدوج و الگوریتم ژنتیک

جدول ۳-۲: تعداد یال‌های برشی گراف توسط الگوریتم گرادیان مزدوج و ژنتیک

جدول ۳-۱: هزینه تابع هدف الگوریتم‌های گرادیان مزدوج و ژنتیک

می‌رود، این نتایج مؤید نتایج جدول قبلی هستند. مجدداً در اینجا هم نتایج هر سطر میانگین ۲۰ اجرای مختلف هستند. بهبود حاصله در نتیجه استفاده از الگوریتم ژنتیک برای پیدا کردن مقادیر اولیه مناسب روش گرادیان مزدوج، زمان پردازش را بیشتر کرده است. هر سطر جدول ۳-۳ زمان موردنیاز برای تمام ۲۰ نمونه تصادفی از هر اندازه گراف برای دو روش را نشان می‌دهد.

در روش افزابندی طیفی گراف برای مجموعه داده‌های بزرگ، محاسبه ماتریس مجاورت داده‌ها و همچنین مقادیر ویژه و بردارهای ویژه ماتریس لاپلاسیان متناظر با آن نیازمند صرف زمان محاسباتی و میزان حافظه ذخیره‌سازی بالا است. روش نیستروم این مشکل را برطرف می‌کند و می‌توان به کمک این روش از افزابندی طیفی برای مجموعه داده‌های بزرگ استفاده کرد.

از موارد استفاده از افزابندی طیفی گراف می‌توان به قطعه بندی تصویر^۱، تجزیه و تحلیل شبکه‌های اجتماعی و خوشه بندی داده‌ها اشاره کرد. در قطعه بندی تصویر، گرافی متناظر با تصویر ساخته می‌شود که هر پیکسل^۲ تصویر به عنوان یک راس در این گراف و اختلاف مولفه‌های رنگی و موقعیت مکانی هر دو پیکسل به عنوان وزن یال متصل کننده آن‌ها در نظر گرفته می‌شود. با تبدیل یک تصویر کوچک 50×50 پیکسل به گراف، تعداد گره‌ها برابر ۲۵۰۰ و ماتریس مجاورت آن $W \in \mathbb{R}^{2500 \times 2500}$ می‌شود. با بزرگ شدن تصویر، تعداد مولفه‌های این ماتریس به اندازه توان دوم پیکسل‌ها افزایش می‌یابد که به سرعت از ظرفیت حافظه برای ذخیره‌سازی خارج می‌شود و پیش از رسیدن به مرحله افزابندی طیفی، متوقف می‌شویم. با کمک روش نیستروم و تقریبی از ماتریس مجاورت با تعدادی از نمونه داده، افزابندی طیفی را بر روی این تصاویر انجام می‌دهیم.

^۱Image Segmentation

^۲Pixel

۲-۳ پیاده‌سازی روش نیستروم بر روی مجموعه داده‌های آیریس و

سرطان سینه

مجموعه داده‌های آیریس^۱ و سرطان سینه^۲ از مجموعه داده‌های UCI انتخاب شده‌اند و از طریق تارنمای آن^۳ قابل دسترسی اند.

مجموعه داده آیریس شامل سه کلاس است که در هر کدام ۵۰ نمونه داده وجود دارد. هر کلاس، از دو کلاس دیگر به صورت خطی جداپذیر است ولی دو به دو، خطی جداپذیر نیستند. نمونه داده‌ها دارای چهار ویژگی اند که به ترتیب طول ساقه، عرض ساقه، طول گلبرگ و عرض گلبرگ به سانتی متر می‌باشند. کلاس‌ها به ترتیب مربوط به گل‌های آیریس ستوسا، وریکولور و ویرجینیکا می‌باشند.

در مجموعه داده سرطان سینه، ویژگی‌ها از یک تصویر دیجیتالی از بافت برداری آسپیراسیون^۴ پستان محاسبه شده‌اند که ویژگی‌های هسته سلولی موجود در تصویر را توصیف می‌کنند. تعداد داده‌ها ۵۶۹ و تعداد ویژگی‌ها برابر ۱۰ تا است. داده‌ها به دو کلاس بدخیم^۵ و خوش‌خیم^۶ تخصیص داده شده‌اند که ۲۱۲ داده به کلاس اول و ۳۵۷ داده به کلاس دوم تعلق دارند.

همانگونه که در فصل قبل گفته شد، در روش نیستروم انتخاب نمونه‌ها در بازدهی الگوریتم تاثیر زیادی دارد. ما انتخاب نمونه‌ها را به دو صورت تصادفی (RS) و بر اساس تابع احتمال افزایشی الگوریتم ۲-۴ (IS) انجام دادیم. هر چند انتخاب این نمونه‌ها به روش تابع احتمال افزایشی بر زمان پردازش الگوریتم می‌افزاید، اما دقت افزاینده را به میزان خوبی افزایش می‌دهد. نتایج این آزمایشات در جداول ۳-۴ تا ۳-۹ نمایش داده شده‌اند. در هر اجرای این الگوریتم، تعداد نمونه داده‌ها را تغییر دادیم و هر سطر این جداول مربوط به تعداد نمونه داده خاصی است. در نمونه‌گیری تصادفی (RS)، الگوریتم را چند بار اجرا کردیم تا نمونه داده‌های تصادفی متفاوتی ایجاد شوند و برای نمونه‌گیری بر پایه احتمال (IS) تعداد مراحل نمونه‌گیری را تغییر دادیم. به دلیل ثابت بودن ماتریس مجاورت گراف متناظر با مجموعه داده‌ها، در فرمول ۲-۵۳ قضیه ۲-۱۰، برای t یکسان، مجموعه S ثابت است. با تغییر t ماتریس E نیز تغییر می‌کند و در نتیجه مجموعه S متفاوتی خواهیم داشت.

در سطر اول برای مجموعه داده آیریس تعداد ۴۰ تا از داده‌ها را به عنوان نمونه داده انتخاب کردیم و به ترتیب سطرهای بعدی مربوط به ۸۰، ۱۰۰، ۱۲۰ و ۱۴۰ نمونه داده می‌باشند. برای هر سطر، نمونه‌گیری تصادفی (RS) را ۳ بار تکرار کرده و نمونه‌های (IS) را به صورت ۵ درصد، ۱۰ درصد و ۲۵ درصد انتخاب کردیم. به این معنی که تعداد مراحل نمونه‌گیری را به ترتیب ۲۰، ۱۰ و ۴ بار تکرار کردیم. جدول ۳-۶ مربوط به زمان اجرای الگوریتم است، همانگونه که انتظار می‌رفت زمان روش (IS) به دلیل پیمایش ماتریس مجاورت در چند مرحله، از روش (RS) بیشتر است. در جدول ۳-۴ میانگین خطا نمایش داده شده است که در بیشتر موارد روش نمونه‌گیری بر پایه احتمال خطای کمتری داشته است و در مجموع نیز خطا را کاهش داده است. جدول ۳-۵ کمترین خطای دو روش با تعداد نمونه داده برابر را نمایش می‌دهد که نشان دهنده

^۱Iris ^۲Wisconsin Diagnostic Breast Cancer ^۳<https://archive.ics.uci.edu/ml/datasets.html> ^۴Fine Needle Aspirate (FNA) ^۵Malignant ^۶Benign

روش		شماره
IS	RS	
۰/۴۹۴	۰/۱۶۵	۱
۰/۴۱۹	۰/۱۰۷	۲
۰/۴۰۶	۰/۱۰۱	۳
۰/۴۳۸	۰/۱۲۲	۴
۰/۴۳۹	۰/۱۵۲	۵
۰/۴۳۹	۰/۱۲۹	میانگین

جدول ۳-۶: زمان اجرای الگوریتم نیستروم برای آیریس

روش		شماره
IS	RS	
۸/۶۶۷	۱۰/۶۶۷	۱
۱۲/۰۰۰	۱۱/۳۳۳	۲
۱۱/۳۳۳	۱۱/۳۳۳	۳
۱۰/۶۶۷	۱۱/۳۳۳	۴
۱۱/۳۳۳	۱۱/۳۳۳	۵
۱۰/۸۰۰	۱۱/۲۰۰	میانگین

جدول ۳-۵: بهترین درصد خطای الگوریتم نیستروم برای آیریس

روش		شماره
IS	RS	
۱۱/۱۱۱	۱۱/۳۳۳	۱
۱۲/۲۲۲	۲۴/۲۲۲	۲
۱۱/۷۷۸	۱۱/۳۳۳	۳
۱۱/۵۵۶	۱۲/۰۰۰	۴
۱۱/۳۳۳	۱۱/۳۳۳	۵
۱۱/۶۰۰	۱۴/۰۴۴	میانگین

جدول ۳-۴: درصد خطای الگوریتم نیستروم برای آیریس

روش		شماره
IS	RS	
۴/۴۴۰	۰/۴۸۸	۱
۴/۳۹۲	۰/۱۳۴	۲
۴/۳۲۷	۰/۱۱۸	۳
۴/۶۲۳	۰/۲۹۴	۴
۴/۴۴۶	۰/۲۵۸	میانگین

جدول ۳-۹: زمان اجرای الگوریتم نیستروم برای سرطان سینه

روش		شماره
IS	RS	
۳/۳۶۷	۲/۷۸۲	۱
۲/۹۲۸	۳/۲۲۱	۲
۲/۴۸۹	۲/۴۸۹	۳
۲/۳۴۳	۱۷/۵۷۰	۴
۲/۷۸۲	۶/۵۱۵	میانگین

جدول ۳-۸: بهترین درصد خطای الگوریتم نیستروم برای سرطان سینه

روش		شماره
IS	RS	
۴/۱۹۷	۱۰/۴۹۳	۱
۷/۱۲۵	۲۰/۰۵۹	۲
۲/۶۳۵	۱۸/۶۴۳	۳
۱۴/۴۴۶	۳۱/۶۲۵	۴
۷/۱۰۱	۲۰/۲۰۵	میانگین

جدول ۳-۷: درصد خطای الگوریتم نیستروم برای سرطان سینه

اهمیت انتخاب نمونه‌ها و تاثیرگذاری در کیفیت خوشه‌بندی است. تمام این مراحل را برای مجموعه داده سرطان سینه نیز انجام دادیم و تعداد نمونه‌ها را به ترتیب برابر با ۶۰، ۸۰، ۱۰۰ و ۱۸۰ گرفتیم و روش (RS) را ۳ بار برای هر نمونه داده تکرار کردیم و این نمونه‌ها را برای روش (IS) به صورت ۱۰ درصد، ۲۵ درصد و ۵۰ درصد انتخاب کردیم. نتایج به دست آمده برای این مجموعه داده در جداول ۳-۷، ۳-۸ و ۳-۹ نشان داده شده‌اند که همانند مجموعه داده آیریس نشان از بهبودی نتایج توسط روش (IS) دارند.

۳-۳ موازی سازی الگوریتم بروزرسانی احتمال

برای محاسبه نقاط نمونه به روش تابع احتمال افزایشی در الگوریتم بروزرسانی احتمال ۲-۵ می‌توان برای بهبود زمان اجرای الگوریتم، یک موازی سازی ساده انجام داد. هنگام بروزرسانی احتمال انتخاب سطرهای ماتریس مجاورت در حلقه سطر (۳) این الگوریتم، احتمال هر سطر مستقل از سطرهای دیگر است و می‌توان بدون تغییر در اصل موضوع حلقه‌ی محاسبه این احتمالات را به صورت موازی اجرا نمود. لازم به ذکر است که برای هر بار اجرای الگوریتم می‌توان این موازی سازی را انجام داد اما به دلیل تغییر و افزایش تعداد نمونه‌ها در بردار S و تغییر در ماتریس E سطر (۲) این الگوریتم نمی‌توان بر روی کل مراحل انتخاب سطرها موازی سازی انجام داد و نتایج به دست آمده تغییر خواهند کرد.

فهرست منابع

- [۱] گریمالدی، رالف پی. ریاضیات گسسته و ترکیبیاتی. ترجمه‌ی عالم‌زاده، علی‌اکبر. نشر علوم نوین، ویرایش چهارم، ۱۳۸۰.
- [2] Golub, G.H. and Van Loan, C.F. *Matrix Computations*. Johns Hopkins University Press, 2nd ed. , 1989.
- [3] M.Apostol, Tom. *Multi Variable Calculus and Linear Algebra with Applications to Differential Equations and Probability*, vol. 2. Indiana University, 2nd ed. , 1968.
- [۴] کاردل، مرتضی. رفاهی شیخانی، امیرحسین. روش مستقل از معکوس وزن دار برای حل معادله مقدار ویژه تعمیم یافته. مجله ریاضی کاربردی واحد لاهیجان، ۲۰۱۰.
- [۵] کارالامبوس، دی. اون‌بورکین، شاو. اصول آنالیز حقیقی. ترجمه‌ی عالم‌زاده، علی‌اکبر. پارتیان، ۱۳۸۹.
- [۶] بوردون، ریچارد. فیرز، داگلاس. آنالیز عددی. ترجمه‌ی عالم‌زاده، علی‌اکبر. بلبلیان، اسماعیل. ققنوس، ویرایش اول، ۱۳۹۲.
- [7] Frieze, A, Kannan, R, and Vempala, S. Fast monte carlo algorithms for finding low-rank approximations. *ACM*, 51:370–378, 1998.
- [8] Mohar, Bojan. The laplacian spectrum of graphs. in *Graph Theory, Combinatorics, and Applications*, pp. 871–898. Wiley, 1991.
- [9] Chung, F.R.K. *Spectral Graph Theory*. American Mathematical Society, 1997.
- [10] Luxburg, Ulrike. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, December 2007.
- [11] Kabelíková, Pavla. Graph partitioning using spectral methods. tech. rep., VŠB - Technical University of Ostrava Faculty of Electrical Engineering and Computer Science Department of Computer Science, 2006.
- [۱۲] راثو، اس. اس. بهینه‌سازی (تئوری و کاربردی). ترجمه‌ی شهیدی‌پور، سیدمحمد مهدی. انتشارات دانشگاه فردوسی، ویرایش اول، ۱۳۷۳.
- [13] MacQueen, J. B. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1:281–297, 1967.

- [14] Hastie, T, Tibshirani R and Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, 2009.
- [15] Shi, Jianbo and Malik, Jitendra. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, August 2000.
- [۱۶] طیبه، فیاض،. خوشه بندی طیفی برای قطعه بندی تصویر. پایان نامه کارشناسی ارشد، دانشکده ریاضی و علوم کامپیوتر، دانشگاه حکیم سبزواری، ۱۳۹۴.
- [۱۷] احمد، ابویی مهریزی،. استفاده از بهینه سازی پیوسته در حل رده ای از مسایل افراز بندی گراف. پایان نامه کارشناسی ارشد، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد، ۱۳۹۴.
- [18] Press, W.H, Flannery B.P and Teukolsky, S.A. *Numerical Recipes in C*. Cambridge University, 1998.
- [19] Kruyt, N.P. A conjugate gradient method for the spectral partitioning of graphs. *Parallel Computing*, 22(11):1493–1502, 1997.
- [20] Potchen, A, Simon H.D and Liou, K.P. Partitioning sparse matrices with eigenvectors of graphs,. *SIAM*, 11:430–452, 1990.
- [21] Hongjie, Jia, Shifei Ding Mingjing Du. A nyström spectral clustering algorithm based on probability incremental sampling. *Soft Comput*, 21(19):5815–5827, October 2017.
- [22] Fowlkes, C, Belongie S Chung F Malik J. Spectral grouping using the nyström method. *IEEE Trans Pattern Anal Mach Intell*, 26(2):214–225, 2004.
- [23] Baker, C.G, Gallivan K.A Dooren P.V. Low-rank incremental methods for computing dominant singular subspaces. *Linear Algebra*, 436(8):2866–2888, 2012.

پیوست آ

اثبات نمونه‌گیری افزایشی مبتنی بر احتمال

برای یک زیر فضای $V \subseteq \mathbb{R}^n$ ، $\pi_{V,k}(A)$ نشان‌دهنده بهترین تقریب مرتبه- k برای A (تحت نرم فرینبوس) با سطرهای V است. در این صورت

$$\pi_k(A) = \pi_{\mathbb{R}^n,k}(A) = \sum_{i=1}^k \sigma_i u^{(i)} v^{(i)T}$$

بهترین تقریب مرتبه- k برای A و $\pi_V(A) = \pi_{V,n}(A)$ تصویر متعامد A بر V است. می‌گوییم مجموعه‌ای از سطرهای A ، که به معنی گروهی از اندیس سطرها، بجای خود سطرهاست. برای مجموعه‌ای از سطرهای A مجموعه S ، می‌گیریم $\text{span}(S) \subseteq \mathbb{R}^n$ نشان‌دهنده زیرفضای تولید شده با این سطرهاست. برای اختصار بجای $\pi_{\text{span}(S)}(A)$ می‌نویسیم $\pi_S(A)$ ، که $\pi_{S,k}(A)$ کوتاه شده $\pi_{\text{span}(S),k}(A)$ است.

برای π_V داریم:

۱. π_V خطی است، به این معنی که $\pi_V(\lambda A + B) = \lambda \pi_V(A) + \pi_V(B)$ برای هر $\lambda \in \mathbb{R}$ و ماتریس‌های $A, B \in \mathbb{R}^{m \times n}$.

۲. اگر $V, W \in \mathbb{R}^n$ زیر فضای خطی متعامد باشد، آنگاه $\pi_{V+W}(A) = \pi_V(A) + \pi_W(A)$ برای هر ماتریس $A \in \mathbb{R}^{m \times n}$ بدست می‌آید.

برای یک بردار تصادفی v احتمال آن با بردار $E(v)$ تعریف شده است، که در آن هر مولفه، مقدار احتمال مولفه‌ای در v است.

برای سهولت اثبات، یک لم را به صورت زیر تعریف می‌کنیم.

لم آ-۱. برای $A \in \mathbb{R}^{m \times n}$ ، با توجه به زیرفضای برداری $V \subseteq \mathbb{R}^n$ ، فرض کنید $E = A - \pi_V(A)$. S یک مجموعه نمونه‌گیری تصادفی است که شامل s سطر انتخاب شده از A است و احتمال انتخاب با توزیع D زیر مطابقت

دارد:

$$P_i = \frac{\|E^{(i)}\|_F^2}{\|E\|_F^2} \quad (1-\bar{A})$$

سپس، برای هر عدد صحیح غیر صفر k ،

$$E_S \left(\|A - \pi_{V+\text{span}(S),k}(A)\|_F^2 \right) \leq \|A - \pi_k(A)\|_F^2 + \frac{k}{s} \|E\|_F^2$$

برهان. بردار

$$\omega^{(1)}, \dots, \omega^{(k)} \in V + \text{span}(S)$$

را در نظر بگیرید، فرض کنید

$$W = \text{span}\{\omega^{(1)}, \dots, \omega^{(k)}\}$$

W می‌تواند به عنوان یک تقریب خوب برای $\text{span}\{v^{(1)}, \dots, v^{(k)}\}$ در نظر گرفته شود، یعنی

$$E_S \left(\|A - \pi_W(A)\|_F^2 \right) \leq \|A - \pi_k(A)\|_F^2 + \frac{k}{s} \|E\|_F^2 \quad (2-\bar{A})$$

جایی که

$$\pi_k(A) = \pi_{\text{span}\{v^{(1)}, \dots, v^{(k)}\}}(A)$$

$\text{span}\{v^{(1)}, \dots, v^{(k)}\}$ نشان‌دهنده بهترین زیرفضای یک تصویر است. همچنان که

$$W \subseteq V + \text{span}(S)$$

اگر شواهدی از فرمول (2- \bar{A}) را ارائه دهیم، لم 1- \bar{A} ثابت خواهد شد.

متغیر تصادفی $X_l^{(j)}$ را تعریف می‌کنیم که احتمال آن P_i است، برای هر $i = 1, \dots, m$ و $l = 1, \dots, s$ وجود دارد

$$X_l^{(j)} = \frac{u_i^{(j)}}{P_i} E^{(i)} = \frac{u_i^{(j)}}{P_i} \left(A^{(i)} - \pi_V(A^{(i)}) \right)$$

$X_l^{(j)}$ یک تابع خطی از یک سطر مشخص از A است که از توزیع D نمونه برداری می‌شود. فرض کنید

$$X^{(j)} = \frac{1}{s} \sum_{l=1}^s X_l^{(j)}$$

سپس

$$E_S(X^{(j)}) = E^T u^{(j)}$$

تعریف می‌کنیم

$$\omega^{(j)} = \pi_V(A)^T u^{(j)} + X^{(j)}, \quad 1 \leq j \leq k \quad (3-\bar{A})$$

سپس به دست می‌آوریم

$$E_S(\omega^{(j)}) = \sigma_j v^{(j)}$$

گام بعدی یافتن central moment مرتبه دوم $\omega^{(j)}$ است، یعنی

$$E_S(\|\omega^{(j)} - \sigma_j v^{(j)}\|^2)$$

همانگونه که

$$\omega^{(j)} - \sigma_j v^{(j)} = X^{(j)} - E^T u^{(j)}$$

بنابراین

$$\begin{aligned} E_S(\|\omega^{(j)} - \sigma_j v^{(j)}\|^2) &= E_S(\|X^{(j)} - E^T u^{(j)}\|^2) \\ &= E_S(\|X^{(j)}\|^2) - 2E_S(X^{(j)} \cdot E^T u^{(j)}) + \|E^T u^{(j)}\|^2 \\ &= E_S(\|X^{(j)}\|^2) - \|E^T u^{(j)}\|^2 \end{aligned} \quad (4-\bar{A})$$

اولین عبارت فرمول (4-آ) را به صورت جداگانه بررسی می‌کنیم:

$$\begin{aligned} E_S(\|X^{(j)}\|^2) &= E_S\left(\left\|\frac{1}{s} \sum_{l=1}^s X_l^{(j)}\right\|^2\right) \\ &= \frac{1}{s^2} \sum_{l=1}^s E_S(\|X_l^{(j)}\|^2) + \frac{2}{s^2} \sum_{1 \leq l_1 < l_2 \leq s} E_S(X_{l_1}^{(j)} \cdot X_{l_2}^{(j)}) \\ &= \frac{1}{s^2} \sum_{l=1}^s E_S(\|X_l^{(j)}\|^2) + \frac{s-1}{s} \|E^T u^{(j)}\|^2 \end{aligned} \quad (5-\bar{A})$$

در فرمول (۵-آ)، $X_{l_1}^{(j)}$ و $X_{l_2}^{(j)}$ از یکدیگر مستقل اند. از فرمول (۴-آ) و فرمول (۵-آ)، نتیجه می‌شود

$$E_S \left(\|\omega^{(j)} - \sigma_j v^{(j)}\|^2 \right) = \frac{1}{s^2} \sum_{l=1}^s E_S \left(\|X_l^{(j)}\|^2 \right) - \frac{1}{s} \|E^T u^{(j)}\|^2 \quad (۶-آ)$$

با توجه به تعریف P_i ، داریم

$$E_S \left(\|X_l^{(j)}\|^2 \right) = \sum_{i=1}^m P_i \frac{\|u_i^j E^{(j)}\|^2}{P_i^2} \leq \|E\|_F^2 \quad (۷-آ)$$

از فرمول‌های (۶-آ) و (۷-آ) یک محدودیت central moment مرتبه دوم از $\omega^{(j)}$ به دست می‌آید:

$$E_S \left(\|\omega^{(j)} - \sigma_j v^{(j)}\|^2 \right) \leq \frac{1}{s} \|E\|_F^2 \quad (۸-آ)$$

با استفاده از این محدودیت می‌توان اثبات را تکمیل کرد. فرض کنید

$$y^{(j)} = \omega^{(j)} / \sigma_j, \quad 1 \leq j \leq k$$

ماتریس

$$F = A \sum_{i=1}^k v^{(i)} y^{(i)T}$$

را تعریف می‌کنیم؛ F می‌تواند خطای $\|A - \pi_W(A)\|_F^2$ را محدود کند. فضای سطری F در عبارت

$$W = \text{span}\{\omega^{(1)}, \dots, \omega^{(k)}\}$$

شامل می‌شود؛ بنابراین $\|A - \pi_W(A)\|_F^2 \leq \|A - F\|_F^2$.

با تجزیه $A - F$ در امتداد بردار منفرد چپ $u^{(1)}, \dots, u^{(r)}$ ، می‌توان از نامساوی (۸-آ) به عنوان کران $\|A - F\|_F^2$

استفاده کرد، بنابراین لم آ-۱ اثبات می شود:

$$\begin{aligned}
 E_S \left(\|A - \pi_W(A)\|_F^{\vee} \right) &\leq E_S \left(\|A - F\|_F^{\vee} \right) \\
 &= \sum_{i=1}^r E_S \left(\left\| (A - f)^T u^{(i)} \right\|_F^{\vee} \right) \\
 &= \sum_{i=1}^k E_S \left(\left\| \sigma_i v^{(i)} - \omega^{(i)} \right\|_F^{\vee} \right) + \sum_{i=k+1}^r \sigma_i^{\vee} \\
 &\leq \frac{k}{s} \|E\|_F^{\vee} + \|A - \pi_k(A)\|_F^{\vee} \tag{۹-آ}
 \end{aligned}$$

□ با کمک لم آ-۱، در ادامه از روش استنتاجی برای اثبات قضیه ۲-۱۰-۲ استفاده می شود.

برهان. فرض کنید t بار نمونه برداری انجام شود. قضیه ۲-۱۰-۲ وضعیت اولیه را وقتی $t = 1$ است، ارائه می دهد، سپس اثبات استنتاجی را شروع می کنیم. فرض کنید

$$E = A - \pi_{S_1 \cup \dots \cup S_{t-1}}(A)$$

با لم آ-۱، برای $s \geq k/\varepsilon$ ، می توان نامساوی زیر را به دست آورد

$$E_{S_t} \left(\|A - \pi_{S_1 \cup \dots \cup S_{t-1} \cup S_t, k}(A)\|_F^{\vee} \right) \leq \|A - \pi_k(A)\|_F^{\vee} + \varepsilon \|E\|_F^{\vee}$$

داریم

$$As \|E\|_F^{\vee} \leq \|A - \pi_{S_1 \cup \dots \cup S_{t-1}, k}(A)\|_F^{\vee}$$

$$E_{S_t} \left(\|A - \pi_{S_1 \cup \dots \cup S_{t-1} \cup S_t, k}(A)\|_F^{\vee} \right) \leq \|A - \pi_k(A)\|_F^{\vee} + \varepsilon \|A - \pi_{S_1 \cup \dots \cup S_{t-1}, k}(A)\|_F^{\vee} \tag{۱۰-آ}$$

با $t - 1$ بار فرض استقرا از نمونه برداری و مقادیر بدست آمده S_1, \dots, S_{t-1} ، می توان نتایج قضیه ۲-۱۰-۲ را استنباط کرد:

$$\begin{aligned}
 E_S \left(\|A - \pi_{S_1 \cup \dots \cup S_t, k}(A)\|_F^{\vee} \right) &\leq \|A - \pi_k(A)\|_F^{\vee} + \varepsilon E_{S_1, \dots, S_{t-1}} \left(\|A - \pi_{S_1 \cup \dots \cup S_{t-1}, k}(A)\|_F^{\vee} \right) \\
 &\leq \|A - \pi_k(A)\|_F^{\vee} + \varepsilon \left(\frac{1}{1 - \varepsilon} \|A - \pi_k(A)\|_F^{\vee} + \varepsilon^{t-1} \|A\|_F^{\vee} \right) \\
 &= \frac{1}{1 - \varepsilon} \|A - \pi_k(A)\|_F^{\vee} + \varepsilon^T \|A\|_F^{\vee} \tag{۱۱-آ}
 \end{aligned}$$



پیوست ب

کد برنامه‌های روش گرادیان مزدوج و روش

نیستروم

برنامه ب-۱: کد گرادیان مزدوج طیفی

```
۱ function [f_value, x] = Spectral_Conjugate(L, jaygasht)
۲
۳ %L is laplacian matrix
۴ %x_k is estimation of eigenvector corresponding to second-
   smallest eigenvalue
۵
۶ format long;
۷ K = size(L,1);
۸ e = ones(1,K);
۹ x = zeros(K,1);
۱۰ for i=1:K
۱۱     x(i)=i-(K+1)/2;
۱۲ end
۱۳
۱۴ switch nargin
۱۵     case 2
۱۶         x=x(jaygasht);
۱۷     case 1
۱۸         x=x(randperm(K));
۱۹ end
۲۰
۲۱ x_k = x-((e*x/(e*e'))*e)';
۲۲ rho_k = x_k'*x_k;
۲۳ y_k = L*x_k;
۲۴ F_k = x_k'*y_k/rho_k;
```

```

۲۵ g_k = 2/rho_k*(F_k*x_k-y_k);
۲۶ h_k = g_k;
۲۷ beta_k = g_k'*g_k;
۲۸
۲۹ for k=1:K
۳۰
۳۱     F = F_k;
۳۲     x = x_k;
۳۳     rho = rho_k;
۳۴     g = g_k;
۳۵     beta = beta_k;
۳۶     h = h_k;
۳۷
۳۸     y_k = L*h;
۳۹     p_k = rho; q_k = 2*x'*h; r_k = h'*h;
۴۰     s_k = rho*F; t_k = 2*x'*y_k; u_k = h'*y_k;
۴۱     alpha_k = (-(u_k*p_k-s_k*r_k)+...
۴۲               sqrt((u_k*p_k-s_k*r_k)^2-(u_k*q_k-t_k*r_k)*(t_k*
                    p_k-s_k*q_k))...
۴۳               )/(u_k*q_k-t_k*r_k);
۴۴     F_k = (s_k+t_k*alpha_k+u_k*alpha_k^2)/(p_k+q_k*alpha_k+r_k*
            alpha_k^2);
۴۵     x_k = x+alpha_k*h;
۴۶
۴۷     if norm(g)<10^-6
۴۸         sprintf('iteration %d',k);
۴۹         break
۵۰     end
۵۱
۵۲     rho_k = x_k'*x_k;
۵۳     g_k = 2/rho_k*(F_k*x_k-F*x+.5*rho*g-alpha_k*y_k);
۵۴     beta_k = g_k'*g_k;
۵۵     h_k = g_k+(beta_k/beta)*h;
۵۶ end
۵۷ x = x_k./norm(x_k);
۵۸ f_value = x'*L*x;

```

برنامه ب-۲: کد خوشه‌بندی به روش نیستروم

```

۱ function labels = Nystrom_Clustering(data, k, nsamp, m)
۲
۳ % k is number of clusters
۴ % m is sampling periods for incremental_sampling
۵ % pdist(a,b) calculate Euclidean distance between vectors a and
   b
۶
۷ n = 25;
۸ t = 0;
۹ niter = 5;

```

```

۱۰ N = size(data,1);
۱۱
۱۲ switch nargin
۱۳     case 4
۱۴         W = pdist(data,data);
۱۵         W = exp(-W.^2/2);% similarity matrix of data
           points
۱۶         ind_rest = 1:N;
۱۷         ind_samp = incremental_sampling(W,nsamp,m);
۱۸         ind_rest(ind_samp) = [];
۱۹         ind = randperm(N-nsamp);
۲۰         ind_rest = ind_rest(ind);
۲۱         ind_all = [ind_samp ind_rest];
۲۲         A = W(ind_samp,ind_samp);
۲۳         B = W(ind_samp,ind_rest);
۲۴     case 3
۲۵         ind_all = randperm(N);
۲۶         ind_samp = ind_all(1:nsamp);
۲۷         ind_rest = ind_all(nsamp+1:end);
۲۸         A = pdist(data(ind_samp,:),data(ind_samp,:));
۲۹         B = pdist(data(ind_samp,:),data(ind_rest,:));
۳۰         A = exp(-A.^2/2);% similarity matrix of the
           sample points
۳۱         B = exp(-B.^2/2);% similarity matrix of the
           sample points and the remaining points
۳۲     otherwise
۳۳         error('The number of entries must be at least 3
           and at most 4');
۳۴ end
۳۵
۳۶ % the joys of nystrom
۳۷ [V,~] = nystrom_ncut(A,B);
۳۸ % depermute the nsytrom result
۳۹ V(ind_all,:) = V;
۴۰ V1 = V(:,1:k); % choose the k eigenvectors
۴۱ Vnormalized = V1./repmat(sqrt(sum(V1.^2,2)),1,k);
۴۲ % clustering k-th eigenvector to k clusters via k-means
۴۳ [~,labels,~] = kmwrapper(Vnormalized',k,n,t,niter);
۴۴ labels=labels';

```

برنامه ب-۳: کد محاسبه بردارهای ویژه یک ماتریس بزرگ به روش نیستروم

```

۱ function [V,ss]=nystrom_ncut(A,B);
۲
۳ Nsamples = size(A,1);
۴ Nothers = size(B,2);
۵ Npix = Nsamples + Nothers;
۶
۷ % compute total connection weight

```

```

۸ disp('computing total connection weight...');
۹ d1 = sum([A;B'],1);
۱۰ d2 = sum(B,1) + sum(B',1)*pinv(A)*B;
۱۱ d = [d1 d2]';
۱۲ disp('done.')
۱۳
۱۴ % normalize
۱۵ disp('normalizing...');
۱۶ v = sqrt(1./d);
۱۷ A = A.*(v(1:Nsamples)*v(1:Nsamples)');
۱۸ B = B.*(v(1:Nsamples)*v(Nsamples+(1:Nothers)))');
۱۹ disp('done.')
۲۰
۲۱ % find eigenvectors via PCA/nystrom trick
۲۲ disp('computing eigenvectors...')
۲۳ [U,S,junk]=svd(A);
۲۴ Asi=U*pinv(sqrt(S))*U';
۲۵ Q=A+Asi*B*B'*Asi;
۲۶ [U,L,junk]=svd(Q);
۲۷ Va=[A;B']*Asi*U*pinv(sqrt(L));
۲۸
۲۹ for i = 1:Nsamples-1
۳۰     V(:,i) = Va(:,i+1)./Va(:,1);
۳۱ end;
۳۲ disp('done.')
۳۳
۳۴ ss=1-diag(L); % convert to (D-W)x=lambda Dx form
۳۵ ss(1)=[]; % drop first eigenvalue (which is 1)

```

برنامه ب-۴: کد نمونه‌گیری احتمالی افزایشی

```

۱ function S = incremental_sampling(W, m, s)
۲
۳ n = length(W);
۴ p = zeros(1,n);
۵ t = m / s;
۶
۷ for i=1:length(p)
۸     p(i) = norm(W(i))^2;
۹ end
۱۰ p = p / norm(W,'fro')^2;
۱۱ [~, sam] = sort(p,'descend');
۱۲ S(1:s) = sam(1:s);
۱۳
۱۴ for i=2:t
۱۵     P = update_probability(W, S);
۱۶     [~, sam] = sort(P,'descend');
۱۷     S(((i-1)*s)+1:i*s) = sam(1:s);
۱۸ end

```

برنامه ب-۵: کد بروزرسانی احتمال

```
۱ function P = update_probability(W, S)
۲
۳ n = length(W);
۴ P = zeros(n,1);
۵ R_perim = zeros(size(W,1),size(W,2));
۶ R_perim(S,:) = W(S,:);
۷ R_perim(:,S) = W(:,S);
۸ [U,~,~] = svd(R_perim);
۹ E = U * W * U';
۱۰ for i=1:n
۱۱     if ~ismember(i,S)
۱۲         P(i) = norm(E(i))^2;
۱۳     end
۱۴ end
۱۵ P = P / norm(P);
```

واژه‌نامه فارسی به انگلیسی

Relaxation	آزادسازی
Validation of the results	اعتبار نتایج
Partitioning	افرازبندی
Genetic Algorithm	الگوریتم ژنتیک
Feature selection	انتخاب ویژگی
Fine Needle Aspirate	بافت برداری آسپیراسیون
Malignant	بدخیم
Eigenvector	بردار ویژه
Graph Cut	برش گراف
Orthonormal Basis	پایه متعامد-یکه
Pixel	پیکسل
Singular Value Decomposition	تجزیه مقدار تکین
Pattern recognition	تشخیص الگو
Computer aided diagnosis	تشخیص به کمک ماشین
Wisconsin Diagnostic Breast Cancer	تشخیص سرطان سینه ویسکانسین
Projection	تصویر شده
Interpretation of the results	تفسیر نتایج
Low-rank Approximation	تقریب مرتبه پایین
Redundant	تکراری
Algebraic Multiplicity	تکرر جبری
Rayleigh quotient	خارج قسمت رایلی
Benign	خوش خیم
Spectral Clustering	خوشه‌بندی طیفی

Rank	رتبه
Nyström Method	روش نیستروم
Rayleigh–Ritz	ریلی-ریتز
Function Space	فضای تابع
Span Space	فضای تولید شده
Fletcher–Reeves	فلچر-ریویس
Image Segmentation	قطعه بندی تصویر
Conjugate Gradient	گرادیان مزدوج
Gradient Descent	گرادیان نزولی
Directed graph	گراف جهت دار
Adjacency Matrix	ماتریس مجاورت
Laplacian Matrix	ماتریس لاپلاسیان
Multi-set	مجموعه چندگانه
Iris dataset	مجموعه داده آیریس
Clustering criterion	محک خوشه بندی
Singular Value	مقدار تکین
Eigenvalue	مقدار ویژه
Proximity measure	میزان نزدیکی
Irrelevant	نامربوط
Multiplicity	نرخ تکرار
Normalization	نرمال سازی
Probability Incremental Sampling	نمونه گیری احتمالی افزایشی
Connected	همبند

واژه‌نامه انگلیسی به فارسی

Adjacency Matrix	ماتریس مجاورت
Algebraic Multiplicity	تکرر جبری
Benign	خوش خیم
Clustering criterion	محک خوشه‌بندی
Conjugate Gradient	گرادیان مزدوج
Connected	همبند
Directed graph	گراف جهت‌دار
Eigenvalue	مقدار ویژه
Eigenvector	بردار ویژه
Feature selection	انتخاب ویژگی
Fine Needle Aspirate	بافت برداری اسپیراسیون
Fletcher-Reeves	فلچر-ریویس
Function Space	فضای تابع
Genetic Algorithm	الگوریتم ژنتیک
Gradient Descent	گرادیان نزولی
Graph Cut	برش گراف
Image Segmentation	قطعه بندی تصویر
Interpretation of the results	تفسیر نتایج
Iris dataset	مجموعه داده آیریس
Irrelevant	نامربوط
Laplacian Matrix	ماتریس لاپلاسیان
Low-rank Approximation	تقریب مرتبه پایین
Malignant	بدخیم

Multiplicity	نرخ تکرار
Multi-set	مجموعه چندگانه
Normalization	نرمال سازی
Nyström Method	روش نیستروم
Orthonormal Basis	پایه متعامد-یکه
Partitioning	افرازبندی
Pixel	پیکسل
Probability Incremental Sampling	نمونه گیری احتمالی افزایشی
Projection	تصویر شده
Proximity measure	میزان نزدیکی
Rank	رتبه
Rayleigh–Ritz	ریلی-ریتز
Rayleigh quotient	خارج قسمت رایلی
Redundant	تکراری
Relaxation	آزادسازی
Singular Value	مقدار تکین
Singular Value Decomposition	تجزیه مقدار تکین
Spectral Clustering	خوشه بندی طیفی
Span Space	فضای تولید شده
Validation of the results	اعتبار نتایج
Wisconsin Diagnostic Breast Cancer	تشخیص سرطان سینه ویسکانسین

Hakim Sabzevari University

An Outline of MSc. Thesis



دانشگاه حکیم سبزواری

Surname:Nemati

Name:Mehdi

Student No.:9413137039

Supervisors: Dr. Mahmoud AminToosi and Dr. Mehdi Zaferanieh

Faculty of Mathematics and Computer Science

Program: Decision Science and Knowledge Engineering Field:

Title of thesis: A Gradient Based Method For The Spectral Graph Partitioning

Keywords:Spectral Graph Partitioning, Conjugate Gradient Method, Nyström Method

Abstract: Data clustering is one of the key issues in the field of data mining and machine learning. One of the clustering methods is to show the dependency of the data by the graph and using graph partitioning methods. Spectral graph partitioning is one of the most famous of these methods that has attracted a lot of attention. In this method, the eigenvector corresponds to the second smallest eigenvalue used for clustering the data. In this research, two approximate methods of Conjugate gradients and Nyström are investigated for calculating eigenvalues and eigenvectors. The conjugate gradient method is a repetitive method that moves from a random start point to the minimum, and the nyström method, using some samples of the data set, obtains an approximate solution to the problem. Finally, both methods have been implemented and improvements have been made on them.



Hakim Sabzevari University

Faculty of Mathematics and Computer Science

**A Thesis Submitted in Partial Fulfilment of the Requirement for the
Degree of Master of Science in Decision Science and Knowledge
Engineering**

A Gradient Based Method For The Spectral Graph Partitioning

Supervisors:

Dr. Mahmoud AminToosi and Dr. Mehdi Zaferanieh

By:

Mehdi Nemati

January 2018