

وزارت علوم، تحقیقات و فناوری
دانشگاه تحصیلات تکمیلی علوم پایه
گاوزنگ - زنجان



استفاده از روش‌های تفکیک منحنی خودمدلساز برای
مطالعه‌ی مزیت مرتبه‌ی دوم در داده‌های دارای همپوشانی
مرتبه

پایان‌نامه کارشناسی ارشد
الناز توکلی

استاد راهنما: دکتر حمید عبداللهی
استاد مشاور: مهسا اکبری

شهریور ۱۳۹۲

چکیده

تعیین کمی آنالیت در حضور مزاحم ناشناسی که در مرحله‌ی کالیبراسیون وارد نشده است یکی از موضوعات مهم در کمومتریکس می‌باشد. در بیشتر موارد، مخلوط‌ها شامل نمونه‌هایی با چند گونه می‌باشند که سیگنال مربوط به آنالیت با مزاحم(ها) همپوشانی دارد. برای رفع این مشکل روش‌های آنالیز چندمتغیره در طی دو دهه‌ی اخیر گسترش پیدا کرده‌اند. در میان این روش‌ها، روش‌های آنالیز چند متغیره-ی بدون فرض اولیه‌ی مدل شیمیایی و فیزیکی، به عنوان روش‌های آنالیز نرم شناخته شده‌اند. داده‌ی مرتبه‌ی دوم به آسانی توسط این روش‌ها آنالیز می‌شود.

به طور کلی دستگاه‌های تجزیه‌ای مدرن در اسپکتروسکوپی، الکتروشیمی و... اطلاعات چند متغیره‌ی مفیدی راجع به یک نمونه در اختیار می‌گذارند که این اطلاعات برای نمونه‌های مختلف در کنار هم قرار می‌گیرند و یک ماتریس داده را که داده‌ی مرتبه‌ی دوم نام دارد، می‌سازند. یک ویژگی ذاتی داده‌ی مرتبه‌ی دوم مزیت مرتبه‌ی دوم نام دارد که اجازه می‌دهد آنالیت حتی در حضور مزاحم ناشناس تعیین شود. این ویژگی حتی با انجام یک کالیبراسیون برای یک نمونه نیز می‌تواند انجام شود و اجازه می‌دهد که آنالیت در حضور هر تعداد مزاحمی به صورت کمی تعیین شود.

راه‌های متفاوتی برای ساخت داده‌ی مرتبه‌ی دوم وجود دارد. در این رساله راهی آسان، کم هزینه و عملی برای ساخت داده‌ی مرتبه‌ی دوم ارائه می‌شود. فرض اصلی این است که فقط یک نمونه حاوی آنالیت و مزاحم‌ها در دست باشد. این نمونه چندین بار و هر بار به درصد مشخصی رقیق می‌شود و از سرهم‌گذاری طیف ثبت شده در هر مرحله، داده‌ی مرتبه‌ی دوم ساخته می‌شود. روش جستجوی شبکه‌ای سیستماتیک مورد استفاده قرار گرفته و با توجه به محدودیت‌هایی که با شناخت از سیستم اعمال شده، یک محدوده از غلظت‌های ممکن و یا پاسخ منحصر به فرد تعیین می‌شود.

در شرایط یکسان، پاسخ آنالیز داده‌ی مرتبه‌ی اول و دوم یکسان است، در بخش دیگر این کار دقت پاسخ-های به دست آمده از آنالیز این داده‌ها با هم مقایسه شد. بررسی‌ها نشان داد که دقت پاسخ حاصل از آنالیز داده‌های مرتبه‌ی دوم به مراتب بیشتر از پاسخ داده‌ی مرتبه‌ی اول می‌باشد.

فهرست

شماره صفحه	فهرست مطالب
۱.....	فصل اول.....
۱.....	مقدمه.....
۱.....	۱.۱ کمومتریکس.....
۲.....	۱.۲ اندازه‌گیری کمی آنالیت در حضور مزاحم ناشناخته در داده‌ی مرتبه‌ی اول.....
۲.....	۱.۲.۱ روش برخورد با اولین صفر.....
۳.....	۱.۲.۲ روش افزایش استاندارد نقطه‌ی H
۴.....	۱.۲.۳ روش تفکیک منحنی نقطه‌ی H
۵.....	۱.۳ اندازه‌گیری کمی آنالیت در حضور مزاحم ناشناخته در داده‌ی مرتبه‌ی دوم.....
۶.....	۱.۴ روش‌های تفکیک منحنی چند متغیره.....
۷.....	۱.۵ روش‌های آنالیز نرم.....
۷.....	۱.۵.۱ روش‌های غیر تکراری.....
۸.....	۱.۵.۲ روش‌های تکراری.....
۸.....	۱.۵.۳ ابهام چرخشی و شدتی.....
۹.....	۱.۶ محدودیت‌ها.....
۱۰.....	۱.۷ تحلیل عاملی.....
۱۰.....	۱.۷.۱ تجزیه مقادیر منفرد (SVD).....
۱۱.....	۱.۸ روش لاوتن و سیلوستر.....
۱۳.....	۱.۹ تفکیک منحنی چند متغیره - حداقل مربعات متناوب (MCR-ALS).....
۱۴.....	۱.۱۰ روش تحلیل مخلوط خودمدل‌ساز ساده با تاثیر متقابل.....
۱۵.....	۱.۱۱ جستجوی شبکه‌ای سیستماتیک از تمامی پاسخ‌های ممکن و نمایش نقشه‌ی خطا در سیستم دو جزئی.....
۱۹.....	۱.۱۲ کمبود مرتبه.....
۲۰.....	۱.۱۲.۱ سرهم زدن.....
۲۱.....	۱.۱۳ ساخت داده‌ی مرتبه‌ی دوم از داده‌ی مرتبه‌ی اول.....
۲۳.....	فصل دوم.....
۲۳.....	مروری بر تحقیقات گذشته.....
۲۴.....	۱.۲ تاریخچه‌ی بخشی از اندازه‌گیرهای کمی آنالیت در حضور مزاحم ناشناخته در داده‌ی مرتبه‌ی اول.....
۲۵.....	۲.۲ بخشی از تاریخچه‌ی استفاده از مزیت داده‌ی مرتبه‌ی دوم.....
۲۷.....	۳.۲ تاریخچه‌ی بخشی از روش‌های محاسبه‌ی محدوده‌ی پاسخ‌های ممکن.....

۲۸	۴ اهداف تحقیق
۲۹	فصل سوم
۲۹	مشابه‌سازی و روش‌های آزمایشگاهی
۲۹	۱. مقدمه
۲۹	۲. منحنی گوسین
۳۰	۳. ۱. شبیه‌سازی و ساخت داده‌ی مرتبه‌ی دوم
	۳. ۳. شبیه‌سازی و بررسی اثر نویز بر دقت داده‌ی مرتبه‌ی دوم ساخته شده به طریق آزمایشگاهی و ساختگی
۳۳	
۳۳	۴. بررسی اثر دقت پاسخ روش‌های مرتبط با داده‌ی مرتبه‌ی اول با روش‌های آنالیز نرم
۳۶	۵. شبیه‌سازی داده برای روش‌های تفکیک منحنی نقطه‌ی H
۳۷	۶. بخش آزمایشگاهی
۳۷	۱. ۶. ۳. دستگاه و نرم‌افزار مورد استفاده
۳۸	۲. ۶. ۳. واکنش‌گرهای مورد استفاده
۳۹	۳. ۶. ۳. روش کار
۴۰	فصل چهارم
۴۰	بحث و نتایج
۴۰	۴. ۱. ساخت داده‌ی مرتبه‌ی دوم
۴۰	۴. ۱. ۱. روش اول
۴۱	۴. ۱. ۲. روش دوم
۴۱	۴. ۲. مزیت مرتبه‌ی دوم در داده‌ی شبیه‌سازی شده
۴۱	۴. ۲. ۱. مثال اول
۴۶	۴. ۲. ۲. مثال دوم
۵۲	۴. ۳. اثر نویز
۵۴	۴. ۴. مقایسه‌ی دقت داده‌ی مرتبه‌ی اول و داده‌ی مرتبه‌ی دوم
۵۷	۴. ۵. نتایج آنالیز با روش تفکیک منحنی نقطه‌ی H
۵۹	۴. ۶. آنالیز داده‌های آزمایشگاهی
۵۹	۴. ۶. ۱. محلول شماره‌ی ۴
۶۱	۴. ۶. ۲. محلول شماره‌ی ۵
۶۴	۴. ۶. ۳. محلول شماره‌ی ۶
۶۶	۴. ۶. ۴. محلول شماره‌ی ۷
۶۷	۴. ۷. نتیجه‌گیری
۶۷	مراجع

فهرست شکل

شماره صفحه

- شکل ۱-۱: شماتیک روش افزایش استاندارد نقطه‌ی **H** ۳
- شکل ۲-۱: تفکیک ماتریس داده به پروفایل‌های غلظتی و طیفی سازنده‌ی آن ۷
- شکل ۳-۱: شماتیک روش لاوتن و سیلوستر ۱۲
- شکل ۴-۱: نمایش نقشه‌ی خطا برای سیستم دو جزئی ۱۸
- شکل ۵-۱: درجه‌های متفاوت ابهام چرخشی ۱۹
- شکل ۶-۱: نمایش روش سرهم زدن ستونی ۲۰
- شکل ۷-۱: نمایش روش سرهم زدن ردیفی ۲۰
- شکل ۸-۱: نمایش روش سرهم زدن ستونی و ردیفی ۲۱
- شکل ۹-۱: شماتیک محدودیت همبستگی ۲۲
- شکل ۱-۳: شکل پروفایل‌های طیفی و غلظتی شبیه‌سازی شده و داده‌ی جذبی ساختگی در مثال اول ۳۱
- شکل ۲-۳: شکل پروفایل‌های طیفی و غلظتی شبیه‌سازی شده و داده‌ی جذبی ساختگی در مثال دوم ۳۲
- شکل ۳-۳: پروفایل طیفی ساختگی گونه‌ها در ۴ مثال برای بررسی دقت پاسخ آنالیز داده‌ی مرتبه‌ی اول و دوم ۳۵
- شکل ۴-۳: پروفایل رقت شبیه‌سازی شده برای بررسی دقت پاسخ آنالیز داده‌ی مرتبه‌ی اول و دوم ۳۵
- شکل ۵-۳: داده‌های جذبی متناظر برای بررسی دقت پاسخ آنالیز داده‌ی مرتبه‌ی اول و دوم ۳۶
- شکل ۶-۳: پروفایل طیفی آنالیت و مزاحم (الف) و داده‌ی جذبی ساختگی (ب) برای بررسی روش تفکیک منحنی نقطه‌ی **H** ۳۷
- شکل ۱-۴: نمایش نقشه‌ی خطای به‌دست آمده توسط روش جستجوی شبکه‌ای در فضای غلظتی (الف) و فضای طیفی (ب) با شرط محدودیت نامنفی بودن برای مثال اول ۴۲
- شکل ۲-۴: نواحی راه‌حل‌های ممکن به‌دست آمده توسط روش جستجوی شبکه‌ای در فضای غلظتی (الف) و فضای طیفی (ب) با شرط محدودیت نامنفی بودن برای مثال اول ۴۲
- شکل ۳-۴: پروفایل‌های غلظتی (الف) و طیفی (ب) ممکن به‌دست آمده توسط روش جستجوی شبکه‌ای برای هریک از گونه‌ها برای مثال اول ۴۳
- شکل ۴-۴: نمایش پروفایل غلظتی محاسبه شده از روش تفکیک منحنی چند متغیره - حداقل مربعات متناوب برای آنالیت، به صورت خط‌چین در مقایسه با تمام پروفایل‌های غلظتی محاسبه شده توسط روش جستجوی شبکه‌ای برای گونه‌ی آنالیت با شرط محدودیت نامنفی بودن ۴۳
- شکل ۵-۴: نواحی راه‌حل‌های ممکن به‌دست آمده توسط روش جستجوی شبکه‌ای در فضای غلظتی (الف) و فضای طیفی (ب) با شرط محدودیت نامنفی بودن و محدودیت منطقه‌ای صفر برای گونه‌ی مزاحم ۴۵
- شکل ۶-۴: نمایش پروفایل غلظتی محاسبه شده به صورت منحصر به فرد از روش جستجوی شبکه‌ای برای گونه‌ی مزاحم با شرط محدودیت‌های نامنفی بودن و محدودیت منطقه‌ای صفر ۴۶
- شکل ۷-۴: نمایش نقشه‌ی خطای به‌دست آمده توسط روش جستجوی شبکه‌ای در فضای غلظتی (الف) و فضای طیفی (ب) با شرط محدودیت نامنفی بودن برای مثال دوم ۴۷

- شکل ۴-۸: نواحی راه‌حل‌های ممکن به‌دست آمده توسط روش جستجوی شبکه‌ای در فضای غلطتی (الف) و فضای طیفی (ب) با شرط محدودیت نامنفی بودن برای مثال دوم ۴۷
- شکل ۴-۹: پروفایل‌های غلطتی (الف) و طیفی (ب) ممکن به‌دست آمده توسط روش جستجوی شبکه‌ای برای هریک از گونه‌ها برای مثال دوم ۴۸
- شکل ۴-۱۰: نمایش پروفایل غلطتی محاسبه شده از روش تفکیک منحنی چند متغیره - حداقل مربعات متناوب برای آنالیت، به صورت خط‌چین در مقایسه با تمام پروفایل‌های غلطتی محاسبه شده توسط روش جستجوی شبکه‌ای برای گونه‌ی آنالیت با شرط محدودیت‌های نامنفی بودن ۴۸
- شکل ۴-۱۱: نواحی راه‌حل‌های ممکن به‌دست آمده توسط روش جستجوی شبکه‌ای در فضای غلطتی (الف) و فضای طیفی (ب) با شرط محدودیت نامنفی بودن و محدودیت منطقه‌ای صفر ۴۹
- شکل ۴-۱۲: نمایش منحنی بار و مشخص کردن ناحیه گزینش‌پذیر بر روی آن برای مثال دوم ۵۰
- شکل ۴-۱۳: پاسخ منحصر به‌فرد محاسبه شده توسط روش جستجوی شبکه‌ای در فضای غلطتی (الف) و فضای طیفی (ب) با شرط محدودیت نامنفی بودن و محدودیت منطقه‌ای صفر و گزینش‌پذیری گونه‌ی آنالیت ۵۱
- شکل ۴-۱۴: نمایش پروفایل غلطتی محاسبه شده از روش تفکیک منحنی چند متغیره - حداقل مربعات متناوب برای آنالیت، به صورت خط‌چین. پروفایل‌های غلطتی محاسبه شده توسط روش جستجوی شبکه‌ای برای گونه‌ی آنالیت بصورت نقطه چین، با شرط محدودیت‌های نامنفی بودن، محدودیت منطقه‌ای صفر و محدودیت گزینش‌پذیری گونه‌ی آنالیت ۵۱
- شکل ۴-۱۵: ناحیه‌ی جواب ممکن در (الف) فضای غلطتی و (ب) فضای طیفی. ناحیه‌ی خط‌چین در حضور محدودیت غیرمنفی بودن، خط راست در حضور محدودیت غیرمنفی بودن و محدودیت منطقه‌ای صفر ۵۲
- شکل ۴-۱۶: مقایسه‌ی ناحیه‌ی ماتریس انتقال **T** برای آنالیت در داده‌ی مرتبه‌ی اول و دوم ۵۵
- شکل ۴-۱۷: لاوتن-سیلوستر پلات برای ۴ مثال شبیه‌سازی شده ۵۶
- شکل ۴-۱۸: تعیین حد بالا و پائین طیف مزاحم در روش تفکیک منحنی نقطه‌ی **H** ۵۸
- شکل ۴-۱۹: طیف مزاحم به‌دست آمده در روش تفکیک منحنی نقطه‌ی **H** ۵۸
- شکل ۴-۲۰: داده‌ی جذبی اندازه‌گیری شده برای محلول شماره‌ی ۴ در حضور سانست زرد به عنوان آنالیت ۶۰
- شکل ۴-۲۱: نمایش بار برای داده‌ی جذبی محلول شماره‌ی ۴ ۶۰
- شکل ۴-۲۲: داده‌ی جذبی اندازه‌گیری شده برای محلول شماره‌ی ۴ در حضور تارترازین به عنوان آنالیت ۶۱
- شکل ۴-۲۳: داده‌ی جذبی اندازه‌گیری شده برای محلول شماره‌ی ۵ در حضور آمارانت به عنوان آنالیت ۶۲
- شکل ۴-۲۴: نمایش بار برای داده‌ی جذبی محلول شماره‌ی ۵ ۶۲
- شکل ۴-۲۵: داده‌ی جذبی اندازه‌گیری شده برای محلول شماره‌ی ۵ در حضور تارترازین به عنوان آنالیت ۶۳
- شکل ۴-۲۶: داده‌ی جذبی اندازه‌گیری شده برای محلول شماره‌ی ۶ در حضور آمارانت به عنوان آنالیت ۶۴
- شکل ۴-۲۷: نمایش بار برای داده‌ی جذبی محلول شماره‌ی ۶ ۶۴
- شکل ۴-۲۸: داده‌ی جذبی اندازه‌گیری شده برای محلول شماره‌ی ۶ در حضور سانست زرد به عنوان آنالیت ۶۵
- شکل ۴-۲۹: داده‌ی جذبی اندازه‌گیری شده برای محلول شماره‌ی ۷ در حضور آمارانت به عنوان آنالیت ۶۶
- شکل ۴-۳۰: نمایش بار برای داده‌ی جذبی محلول شماره‌ی ۷ ۶۷

فهرست جدول

شماره صفحه

جدول ۱-۳: محلول‌های ساخته شده در آزمایشگاه.....	۳۸
جدول ۱-۴: نتایج آنالیز مثال اول در حضور محدودیت نامنفی بودن.....	۴۴
جدول ۲-۴: نتایج آنالیز مثال اول در حضور محدودیت‌های نامنفی بودن و منطقه‌ای صفر.....	۴۶
جدول ۳-۴: نتایج آنالیز مثال دوم در حضور محدودیت نامنفی بودن.....	۴۸
جدول ۴-۴: نتایج آنالیز مثال دوم در حضور محدودیت‌های نامنفی بودن و منطقه‌ای صفر.....	۴۹
جدول ۵-۴: نتایج آنالیز مثال اول در حضور محدودیت‌های نامنفی بودن و منطقه‌ای صفر و گزینش-پذیری.....	۵۱
جدول ۶-۴: نتایج بررسی دقت در داده‌های آزمایشگاهی و ساختگی در مثال اول.....	۵۳
جدول ۷-۴: نتایج بررسی دقت در داده‌های آزمایشگاهی و ساختگی در مثال دوم.....	۵۳
جدول ۸-۴: بررسی دقت داده‌های مرتبه‌ی اول و دوم در مثال الف.....	۵۶
جدول ۹-۴: بررسی دقت داده‌های مرتبه‌ی اول و دوم در مثال ب.....	۵۷
جدول ۱۰-۴: بررسی دقت داده‌های مرتبه‌ی اول و دوم در مثال ج.....	۵۷
جدول ۱۱-۴: بررسی دقت داده‌های مرتبه‌ی اول و دوم در مثال د.....	۵۷
جدول ۱۲-۴: نتایج آنالیز محلول شماره‌ی چهار با فرض سانست زرد به عنوان آنالیت.....	۶۱
جدول ۱۳-۴: نتایج آنالیز محلول شماره‌ی چهار با فرض تارترازین به عنوان آنالیت.....	۶۱
جدول ۱۴-۴: نتایج آنالیز محلول شماره‌ی پنج با فرض آمارانت به عنوان آنالیت.....	۶۳
جدول ۱۵-۴: نتایج آنالیز محلول شماره‌ی پنج با فرض تارترازین به عنوان آنالیت.....	۶۳
جدول ۱۶-۴: نتایج آنالیز محلول شماره‌ی شش با فرض آمارانت به عنوان آنالیت.....	۶۵
جدول ۱۷-۴: نتایج آنالیز محلول شماره‌ی شش با فرض سانست زرد به عنوان آنالیت.....	۶۶
جدول ۱۸-۴: نتایج آنالیز محلول شماره‌ی هفت با فرض آمارانت به عنوان آنالیت.....	۶۷

نمادگذاری

در رساله‌ی حاضر حروف بزرگ و پررنگ مانند \mathbf{A} جهت نمایش ماتریس‌ها (داده‌های مرتبه‌ی دوم)، حروف کوچک و پررنگ مانند \mathbf{a} جهت نمایش بردارها (داده‌های مرتبه‌ی اول) و حروف کوچک مانند a جهت نمایش اعداد (داده‌های مرتبه‌ی صفر) به کار برده شده‌اند. ترانهاده‌ی یک ماتریس یا بردار توسط بالانویس ' \mathbf{T} ' همانند \mathbf{A}^T مشخص شده است. برای ماتریسی مانند \mathbf{A} ، ماتریسی مانند \mathbf{A}^{-1} نشان دهنده‌ی معکوس آن ماتریس می‌باشد.

فصل اول

مقدمه

۱.۱ کمومتریکس

پیشرفت دستگاه‌های اندازه‌گیری در شیمی امکان به‌دست آوردن انبوه داده‌های آزمایشگاهی را فراهم کرد. همزمان شدن این پیشرفت در دهه‌ی ۱۹۷۰ با رشد تکنولوژی کامپیوتر و علمی مانند آمار و ریاضی و امکان کاربرد آن‌ها در شیمی و به‌خصوص شیمی تجزیه، منجر به پیشرفت کمومتریکس شد [۱]. اولین بار یک شیمیدان آلی سوئدی به‌نام سوانت وُلد^۱ در سال ۱۹۷۱ واژه‌ی کمومتریکس^۲ را ابداع کرد. چند سال بعد وُلد همراه با بروس آر. کوالسکی^۳ که یک شیمی تجزیه‌دان آمریکایی بود، انجمن بین‌المللی کمومتریکس^۴ را بنیان نهادند. هر دو افراد از پیشگامان علم کمومتریکس محسوب می‌شوند. تعریفی که توسط این انجمن برای کمومتریکس ارائه شده به این صورت خلاصه می‌شود:

کمومتریکس در موقعیت کنونی یک زیر نظام شیمیایی است که منجر به کاربرد ریاضی و آمار در موارد زیر می‌شود:

-ارزیابی و تفسیر داده‌های شیمیایی و تجزیه‌ای

-بهبودسازی و مدل کردن فرایندها و آزمایش‌های شیمیایی و تجزیه‌ای

-استخراج اطلاعات شیمیایی و تجزیه‌ای از داده‌های تجربی

¹ Svant Wold

² Chemometrics

³ Bruce R. Kowalski

⁴ International Chemometrics Society

یکی از بخش‌های مهم علم کمومتریکس آنالیز داده‌های چند متغیره می‌باشد. استفاده از آنالیز داده‌های چند متغیره نسبت به آنالیز داده‌ی تک متغیره اثر نويز را کمتر کرده و اطلاعات بیشتری وارد سیستم مورد مطالعه می‌کند. [۳ و ۲]

۲.۱ اندازه‌گیری کمی آنالیت در حضور مزاحم ناشناخته^۱ در داده‌ی مرتبه‌ی اول^۲

یکی از جدی‌ترین مشکلات به هنگام اندازه‌گیری کمی آنالیت می‌تواند حضور گونه‌ی مزاحمی باشد که سیگنال آنالیت را تحت تاثیر قرار می‌دهد در حالی که در بخش کالیبراسیون وارد نشده است. اگر داده‌ی به‌دست آمده به ازای هر نمونه، تک متغیره و از مرتبه‌ی صفر^۳ باشد نمی‌توان حضور مزاحم را تشخیص داد و در این صورت استفاده از مدل ساخته شده از سیگنال آنالیت در بخش کالیبراسیون برای پیش‌بینی غلظت آنالیت در نمونه‌ی مجهول اشتباه می‌باشد. اگر داده‌ی به‌دست آمده در آزمایشگاه به ازای هر نمونه، از مرتبه‌ی اول و به‌صورت یک بردار^۴ باشد، می‌توان حضور مزاحم را تشخیص داد. روش‌هایی وجود دارند که با داده‌های مرتبه‌ی اول، غلظت آنالیت را در حضور مزاحم ناشناخته اندازه‌گیری می‌کنند. [۴]

۱.۲.۱ روش برخورد با اولین صفر^۵

برخی از روش‌های آنالیز چند متغیره به‌جای یک پاسخ منحصربه‌فرد برای غلظت یک گونه، یک دامنه‌ایی از غلظت‌های ممکن را محاسبه می‌کنند. اگر داده‌ی اندازه‌گیری شده برداری به‌نام (\mathbf{m}) باشد، این بردار حاصل جمع بردار سیگنال مربوط به مزاحم (\mathbf{b}) و سیگنال آنالیت (\mathbf{s}) در یک غلظت مشخص می‌باشد. در این روش ضریبی به نام (α) تعریف می‌شود که در سیگنال آنالیت ضرب شده و از سیگنال مخلوط کم می‌شود:

$$\mathbf{b} = \mathbf{m} - \alpha \mathbf{s} \quad (1-1)$$

مقدار این ضریب از صفر شروع شده، به مقدار^۶ مشخصی افزایش پیدا می‌کند و هر بار با ضرب شدن این ضریب در سیگنال آنالیت و کم شدن آن از سیگنال نمونه، یک بردار (\mathbf{b}) محاسبه می‌شود. این امر تا جایی ادامه پیدا می‌کند که اولین مقدار صفر در پروفایل باقیمانده (\mathbf{b}) دیده شود. به این- ترتیب یک محدوده غلظتی برای آنالیت محاسبه می‌شود که مقادیر α می‌باشد. محدوده‌ی غلظتی

¹ Unknown Interference

² First order data

³ Zero order data

⁴ Vector

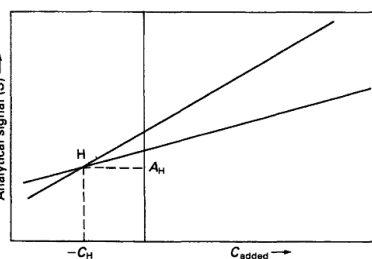
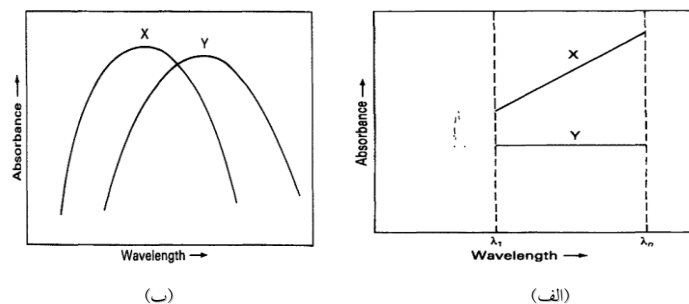
⁵ First zero crossing

⁶ Increment

محاسبه شده از صفر شروع می شود و بالاترین مقدار آن زمانی است که تمام سیگنال کل، مربوط به آنالیت فرض شده باشد و حداقل در یک نقطه، در بردار باقیمانده‌ی (b) که مربوط به مزاحم فرض شده مقادیر منفی ایجاد شود. [5]

۲.۲.۱ روش افزایش استاندارد نقطه‌ی H

این روش در سال ۱۹۸۸ توسط فالکو و همکارانش ارائه شد [۶] و توانایی اندازه‌گیری همزمان دو گونه یا تعیین آنالیت در حضور مزاحم شناخته شده در یک بافت نمونه را داراست. این روش بر مبنای ترکیب مناسبی از سیگنال‌های بدست آمده (برای مثال جذب در اسپکتروفتومتری) برای نمونه در دو متغیر متفاوت (برای مثال دو طول موج متفاوت در اسپکتروفتومتری) می باشد. این دو متغیر طوری انتخاب می شوند که سیگنال مربوط به گونه‌ی مزاحم در این دو متغیر یکسان و سیگنال مربوط به گونه‌ی آنالیت ماکزیمم اختلاف را داشته باشند (شکل ۱-۱ ب). شرط ماکزیمم اختلاف سیگنال مربوط به آنالیت در این دو متغیر انتخابی به دلیل رسیدن به بیشترین حساسیت است. در ادامه افزایش استاندارد آنالیت بر روی نمونه انجام می شود و از ترسیم سیگنال تجزیه‌ای بر حسب غلظت آنالیت اضافه شده در هر یک از این دو متغیر، دو خط راست به دست می آید که این دو خط یکدیگر را در نقطه‌ای به نام H قطع می کنند. طول نقطه‌ی H غلظت آنالیت و عرض نقطه‌ی H سیگنال تجزیه‌ای مربوط به گونه‌ی مزاحم را نشان می دهد (شکل ۱-۱ ج).



(ج)

شکل ۱-۱ شماتیک روش افزایش استاندارد نقطه‌ی H

¹ H.Point Standard Addition Method (HPSAM)

۳.۲.۱ روش تفکیک منحنی نقطه^۱H

مخلوطی شامل دو گونه‌ی آنالیت و مزاحم را در نظر بگیرید به طوری که \mathbf{a} بردار جذب مربوط به آنالیت و \mathbf{b} بردار جذب مربوط به مزاحم (یا مزاحم‌های) ناشناخته باشند. چون طیف آنالیت معلوم و مشخص است، یک طول موج مرجع روی طیف آن انتخاب شده و بقیه محاسبات انجام شده در جهت محاسبه طیف کامل مزاحم خواهد بود.

در هر طول موجی مانند i جذب برای نمونه‌ی مخلوط، برابر جذب گونه‌ی آنالیت \mathbf{a} و گونه‌ی مزاحم \mathbf{b} می‌باشد:

$$\mathbf{s}_i = \mathbf{b}_i + \mathbf{a}_i \quad (۲-۱)$$

که در اینجا \mathbf{s}_i جذب نمونه در طول موج i ام است. زیرنویس i به جذب در یک طول موج معین در محدوده‌ی مورد نظر مربوط می‌شود. \mathbf{a}_i جذب آنالیت و \mathbf{b}_i جذب مزاحم در طول موج i ام است. بعد از انتخاب یک طول موج مرجع از روی طیف آنالیت، \mathbf{k}_i به این صورت تعریف می‌شود:

$$\mathbf{k}_i = \mathbf{a}_{ref} / \mathbf{a}_i \quad (۳-۱)$$

در این رابطه \mathbf{a}_{ref} جذب آنالیت در طول موج مرجع و \mathbf{a}_i جذب آنالیت در هر طول موج می‌باشد. سپس سهم گونه‌ی آنالیت از سیگنال کلی از طریق تفریق $\mathbf{s}_{ref} - \mathbf{k}_i \mathbf{s}_i$ محاسبه شده و این معادله به دست آید:

$$\mathbf{s}_{ref} - \mathbf{k}_i \mathbf{s}_i = \mathbf{a}_{ref} + \mathbf{b}_{ref} - \mathbf{k}_i (\mathbf{b}_i + \mathbf{a}_i) \quad (۴-۱)$$

در رابطه‌ی فوق \mathbf{s}_{ref} جذب مخلوط در طول موج مرجع است. اگر رابطه نوآرایی شود، معادله‌ی زیر ایجاد می‌شود:

$$\mathbf{s}_{ref} - \mathbf{k}_i \mathbf{s}_i = \mathbf{b}_{ref} - \mathbf{k}_i \mathbf{b}_i \quad (۵-۱)$$

رابطه‌ی فوق نشان می‌دهد که نتایج ایجاد شده تنها به گونه‌ی مزاحم وابسته بوده و دیگر گونه‌ی آنالیت در آن‌ها سهمی ندارد. اگر یک‌بار دیگر رابطه نوآرایی شود، معادله‌ی زیر بدست می‌آید:

$$\mathbf{b}_i = [(\mathbf{s}_{ref} - \mathbf{k}_i \mathbf{s}_i) - \mathbf{b}_{ref}] / -\mathbf{k}_i = \mathbf{s}_i - [(\mathbf{s}_{ref} - \mathbf{b}_{ref}) / \mathbf{k}_i] \quad (۶-۱)$$

از آنجا که در رابطه‌ی فوق تمام مقادیر به جز \mathbf{b}_{ref} معلوم و مشخص می‌باشد، تنها با وارد کردن یک مقدار مناسب برای \mathbf{b}_{ref} طیف گونه‌ی مزاحم به دست می‌آید. پس مرحله‌ی بعدی تخمین مقدار \mathbf{b}_{ref} است. در این روش با تعیین یک محدوده برای سیگنال مزاحم در طول موج مرجع، برای آن یک حد بالا و یک حد پایین جذب تعریف می‌شود. حد بالای جذب مزاحم در طول موج مرجع، می‌تواند برابر کل سیگنال نمونه باشد و حد پائین آن کمترین مقداری است که طیف منفی برای مزاحم ایجاد

^۱ H-Point Curve Isolation Method

نکند. طیف واقعی مزاحم می‌تواند یکی از این طیف‌های تعیین شده توسط حد بالا، یا حد پائین یا مابین این‌ها باشد. برای محاسبه‌ی جذب گونه‌ی مزاحم در طول موج مرجع نیاز به یافتن طول موج-هایی می‌باشد که در این طول موج‌ها، جذب برای تمام این طیف‌ها یکسان باشد. برای سهولت در دستیابی به جفت طول موج‌های مناسب، طیف‌هایی را که به آن‌ها بیشترین و کمترین مقادیر b_{ref} تعلق گرفته و بیشترین اختلاف را با یکدیگر دارا می‌باشند، در نظر می‌گیرند. اگر اختلاف جذب در دو طول موج انتخاب شده در طیف‌های اول و آخر تقریباً یکسان باشد، در بقیه طیف‌ها که در این محدوده قرار دارند نیز یکسان خواهد بود. در ضمن برای اطمینان از صحت پیشگویی جذب مزاحم در طول موج مرجع، جذب آنالیت در طول موج‌های انتخاب شده باید تا حد امکان متفاوت باشد. اگر فرض کنیم که دو طول موج انتخاب شده λ_m و λ_n باشند که در آن‌ها $b_m = b_n$ در این صورت یک دستگاه سه معادله و سه مجهول خواهیم داشت:

$$S_{ref} - k_m S_m = b_{ref} - k_m b_m \quad (7-1)$$

$$S_{ref} - k_n S_n = b_{ref} - k_n b_n \quad (8-1)$$

$$b_m = b_n \quad (9-1)$$

که مقدار واقعی b_{ref} از این معادله حاصل می‌شود:

$$b_{ref} = S_{ref} - k_m S_m + [k_m \{ (S_{ref} - k_m S_m) - (S_{ref} - k_n S_n) \} / (k_n - k_m)] \quad (10-1)$$

در این روش، محدوده‌ی ممکن برای طیف گونه مزاحم کاهش می‌یابد و در نهایت با تکرار این مراحل طیف مزاحم، پیش‌بینی می‌شود. سرانجام با انجام کالیبراسیون در این دو طول موج، غلظت آنالیت در نمونه مجهول پیش‌بینی می‌شود [7].

$$-C_H = (A_2 - A_1) / (m_1 - m_2) \quad (11-1)$$

A_1 و A_2 به ترتیب میزان جذب در طول موج اول و دوم، m_1 و m_2 شیب خط کالیبراسیون برای طول موج اول و دوم می‌باشند.

۳.۱ اندازه‌گیری کمی آنالیت در حضور مزاحم ناشناخته در داده‌ی مرتبه‌ی دوم^۱

در داده‌ی مرتبه‌ی دوم که به فرم ماتریس می‌باشد، اندازه‌گیری آنالیت در حضور مزاحم ناشناخته به دلیل اطلاعات بیشتری که وارد سیستم می‌شود، راحت‌تر صورت می‌گیرد و این توانایی، مزیت داده-ی مرتبه دوم^۲ نامیده می‌شود. راه‌های متفاوتی برای به دست آوردن داده‌ی مرتبه‌ی دوم وجود دارد که پاسخ‌های حاصل از آنالیز داده‌های آن‌ها بسیار دقیق و قابل اعتماداند.

¹ Second order data

² Second order advantage

۴.۱ روش‌های تفکیک منحنی چند متغیره^۱

همانطور که گفته شد آنالیز داده‌های چند متغیره یکی از بخش‌های مهم در علم کمومتریکس است. یکی از انواع داده‌های چند متغیره، داده‌ی مرتبه‌ی دوم یا ماتریس داده می‌باشد، طوری که هر ردیف از این ماتریس می‌تواند بیانگر سیگنال ثبت شده برای یک نمونه و هر ستون آن بیانگر تغییرات نمونه‌ها در یک متغیر باشد. روش‌های تفکیک منحنی چند متغیره [۸] روش‌هایی هستند که در آن‌ها هدف، تفکیک ماتریس داده به ماتریس‌های کوچک‌تر می‌باشد که دارای مفاهیم شیمیایی باشند. اگر فرض کنیم ماتریس داده‌ی مورد نظر از روش‌های اسپکتروسکوپی به دست آمده باشد، ماتریس‌های به دست آمده از روش‌های تفکیک منحنی چند متغیره ماتریس پروفایل‌های غلظتی و ماتریس پروفایل‌های طیفی مربوط به گونه‌های موجود در نمونه خواهند بود. زیرا طبق قانون بیر - لامبرت^۲، برای یک تابش تک‌فام، جذب تناسب مستقیم با ضخامت محیط جاذب (b) و غلظت گونه‌ی جاذب (c) دارد.

$$A = \epsilon bc \quad (12-1)$$

معمولاً (b) را ثابت و برابر با 1cm در نظر می‌گیرند. بنابراین می‌توان نوشت $A = \epsilon c$ که در آن، ϵ یک ثابت اختصاصی برای جسم حل شده است که ضریب جذب مولی خوانده می‌شود و مقدار آن وابسته به طول موج نور تابشی می‌باشد. چنانچه از این معادله برمی‌آید اگر نمونه حاوی یک گونه باشد، جذب این محلول از حاصل ضرب ضریب جذب مولی و غلظت گونه به دست می‌آید. برای تعداد گونه‌های بیشتر، قانون جذب بیر جمع‌پذیر بوده و می‌توان نوشت:

$$D = c_1 \epsilon_1 + c_2 \epsilon_2 + \dots + c_n \epsilon_n = c^T s \quad (13-1)$$

بنابراین طبق رابطه‌ی (۱۳-۱) جذب کل یک محلول چند جزئی در یک طول موج خاص از حاصل ضرب برداری غلظت گونه‌ها و ضرایب جذب مولی متناظر با آن‌ها به دست می‌آید. این مفهوم می‌تواند به شکل اندازه‌گیری جذب در یک فرآیند و در گستره‌ای از طول موج‌ها بسط یابد. در این صورت داده‌ی **D** به فرم ماتریس خواهد بود طوری که هر ردیف آن حاصل جمع جذب تمام گونه‌ها در یک زمان و یا برای یک نمونه‌ی خاص در گستره‌ی طول موجی معین می‌باشد و هر ستون آن مقادیر جذب برای نمونه‌های مختلف در یک طول موج مشخص می‌باشد. بنابر قانون بیر می‌توان این داده را به صورت حاصل ضرب پروفایل‌های غلظتی **C** و طیفی **S** خالص اجزاء نوشت.

¹ Multivariate Curve Resoluton (MCR)

² Beer-Lambert Law

$$\begin{matrix} n\lambda \\ \boxed{D} \\ ns \end{matrix} = \begin{matrix} nc \\ \boxed{C} \\ ns \end{matrix} \begin{matrix} n\lambda \\ \boxed{S^T} \\ nc \end{matrix} + \begin{matrix} n\lambda \\ \boxed{R} \\ ns \end{matrix}$$

شکل ۱-۲ تفکیک ماتریس داده به پروفایل‌های غلظتی و طیفی سازنده‌ی آن

در شکل (۱-۲) D ماتریس داده با تعداد ns طیف ثبت شده و تعداد $n\lambda$ طول موجی است که در آن‌ها جذب اندازه‌گیری شده است و nc نیز تعداد اجزای شیمیایی موجود در سیستم است. همچنین ستون‌های ماتریس C پروفایل غلظتی و ردیف‌های ماتریس S^T پروفایل طیفی خالص اجزاء را نشان می‌دهند. همچنین R ماتریسی هم‌بعد ماتریس D می‌باشد و ماتریس باقیمانده نامیده می‌شود که شامل تغییرات تصادفی سیگنال است.

۱.۵ روش‌های آنالیز نرم^۱

چندین روش هدف بالا را میسر می‌کنند که به عنوان روش‌های مدل‌سازی نرم یا روش‌های مستقل از مدل^۲ یا خود مدل‌ساز^۳ شناخته می‌شوند. مشخصه‌ی بارز آن‌ها عدم نیاز به مدل شیمیایی جهت تحلیل و تفکیک ماتریس داده‌هاست. تنها فرض اساسی در این روش‌ها وجود رابطه‌ی دوخطی^۴ در ماتریس داده است. البته با توجه به نوع سیستم و اطلاعات موجود فرضیات دیگری هم در این روش‌ها به کار می‌روند. از مهم‌ترین این فرضیات منفی نبودن پروفایل‌های غلظتی و طیفی اجزاء می‌باشد. محدودیت اصلی همه‌ی روش‌های مدل‌سازی نرم این است که این امکان وجود دارد که حل حاصل از آن‌ها منحصر به فرد نبوده و همراه با ابهام باشد و یک دامنه از جواب‌های ممکن حاصل شود [۹ و ۱۰]. چندین دسته از روش‌های مستقل از مدل وجود دارند که هر یک اهداف متفاوتی را دنبال کرده و همچنین مفاهیم و نتایج متفاوتی دارند. این روش‌ها به دو دسته‌ی عمده تقسیم می‌شوند.

۱.۵.۱ روش‌های غیرتکراری^۵

این روش‌ها روی یافتن یکی از ماتریس‌های اجزای خالص تمرکز دارند. به‌طور مثال در داده‌های جذبی پروفایل‌های غلظتی یا طیفی به‌دست می‌آیند و اغلب الگوریتم یک مرحله‌ای دارند و سریع

¹ Soft modeling

² Model-free analysis

³ Self-modeling

⁴ Bilinear

⁵ Non-iterative methods

می‌باشند [۱۱-۱۶]. این روش‌ها معمولاً تحلیل کاملی از سیستم ارائه نمی‌کنند، بلکه نوعی تخمین در اختیار می‌گذارند. از جمله‌ی این روش‌ها می‌توان به تحلیل عاملی تکاملی^۱ (EFA) [۱۷] اشاره کرد. EFA تنها برای فرآیندهایی که اجزا در آن به‌طور منظم تکامل می‌یابند، قابل کاربرد است. از دیگر مثال‌های این روش‌ها می‌توان از تحلیل مخلوط خودمدل‌ساز ساده با تاثیر متقابل^۲ (SIMPLISMA) [۱۸] نام برد.

۱.۵.۲ روش‌های تکراری^۳

در این روش‌ها هدف، تفکیک کامل ماتریس داده به پروفایل‌های غلظتی و طیفی است. این دسته از روش‌ها با تخمین‌های اولیه‌ای از پروفایل‌های غلظتی و یا طیفی خالص اجزاء شروع به‌کار کرده و طی تکرارهایی آن‌ها را بهینه می‌کنند. بهینه‌سازی تا زمانی ادامه می‌یابد که معیار همگرایی^۴ فراهم شده باشد (تعداد تکرارها میسر شده باشد و یا مجموع مربعات عناصر ماتریس باقیمانده کمتر از یک مقدار خاص باشد). بنابراین این روش‌ها اغلب نسبت به روش‌های غیرتکراری طولانی هستند. برای تخمین اولیه در این روش‌ها، از نتایج روش‌هایی مثل SIMPLISMA، EFA و... و حتی از تخمین‌های تصادفی استفاده می‌کنند.

۱.۵.۳ ابهام چرخشی^۵ و شدتی^۶

تالر^۷ و همکارانش وجود دو نوع ابهام را تشخیص دادند [۱۹]. در یکی از آن‌ها که ابهام چرخشی نامیده می‌شود، ماتریس داده از پروفایل‌هایی ساخته می‌شود که در شکل با پروفایل‌های اصلی متفاوتند اما همگی داده‌ی موردنظر را بازسازی نموده و از معیار همگرایی تعریف شده، تبعیت می‌کنند. در نوع دیگر که ابهام شدتی نام دارد، ماتریس داده از پروفایل‌هایی ساخته می‌شود که از نظر شکل با پروفایل‌های واقعی یکسان‌اند و تنها از نظر شدتی با آن‌ها متفاوتند. چون ضرائب بی‌شماری می‌توان در نظر گرفت که در پروفایل‌های اصلی ضرب شده و داده اصلی را بازسازی نمایند. برای رفع این ابهام می‌توان شدت خاصی را برای پروفایل‌ها تعریف کرد تا امکان تغییر شدت آن‌ها وجود نداشته باشد. در صورت وجود ابهام چرخشی محدوده‌ایی از پاسخ‌های ممکن وجود خواهد داشت که پاسخ واقعی را نیز در بر می‌گیرند. هرچه اطلاعات از سیستم مورد مطالعه بیشتر باشد، محدوده‌ی

¹ Evolving Factor Analysis

² Simple-to-use interactive Self-modeling Mixture Analysis

³ Iterative methods

⁴ Convergence criterion

⁵ Rotational ambiguity

⁶ Intensity ambiguity

⁷ Tauler

پاسخ‌های ممکن نیز کاهش می‌یابد و در برخی موارد پاسخ منحصر به فرد حاصل می‌شود. این اطلاعات به صورت اعمال یکسری محدودیت‌ها در محاسبات وارد می‌شود.

۶.۱ محدودیت‌ها

اگر چه در روش‌های مدل‌سازی نرم نیاز به اطلاعات اولیه‌ای در مورد سیستم شیمیایی تحت مطالعه نیست اما دانش اضافی می‌تواند مبنای کاهش ابهام در تجزیه ماتریس داده و همچنین در نتایج باشد. افزودن این اطلاعات با استفاده از اعمال محدودیت‌ها امکان‌پذیر است [۲۰].

انواع محدودیت‌های متداول:

محدودیت منفی نبودن^۲ [۹]: از آنجایی که پروفایل‌های گونه‌ها در اکثر سیستم‌ها مثبت است، به طور مثال همه‌ی پروفایل‌های غلظتی و طیفی که نتایج جذب و فلورسانس هستند مثبت‌اند، برای همین محدودیتی طراحی شده تا مقادیر منفی در پروفایل‌ها را برابر یا بزرگتر از صفر قرار دهد.

محدودیت تک‌قله‌ای^۳ [۲۰ و ۲۱]: محدودیتی است که حضور فقط یک قله را در پروفایل‌ها اجازه می‌دهد. این شرط در بسیاری از پروفایل‌های پیک شکل همچون کروماتوگرام‌ها یا حتی برای پروفایل‌هایی که یک ماکزیمم ثابت یا یک میل کاهشی دارند به کار می‌رود.

محدودیت بسته بودن^۴ [۸]: محدودیتی برای سیستم‌های بسته است که اصول موازنه جرم بر آن‌ها صادق است. با این محدودیت در هر مرحله از الگوریتم مجموع غلظت همه‌ی اجزا برابر با یک مقدار ثابت قرار می‌گیرد. سیستم بسته مانند تیتراسیون یک اسید که در هر مرحله از واکنش غلظت کل اجزا مقداری ثابت است.

محدودیت برابری^۵ [۸ و ۲۲]: اگر اطلاعات شیمیایی گونه‌ای به صورت خالص وجود داشته باشد، به طور مثال پروفایل طیفی یا غلظتی خالص یک یا چند گونه، در صورت معلوم به صورت یک محدودیت برابری اعمال می‌شود بدین معنا که در طی بهینه‌سازی پروفایل موردنظر بدون تغییر باقی می‌ماند.

محدودیت مدل‌های فیزیکو شیمیایی [۲۳-۲۴]: اگر مدل شیمیایی بخشی از فرایندی که در سیستم اتفاق افتاده در دسترس باشد، این اطلاعات به صورت مدل فیزیکوشیمیایی به عنوان محدودیت در الگوریتم روش‌های مدل‌سازی نرم وارد می‌شود.

¹ Constraint

² Non-negativity

³ Unimodality

⁴ Closure

⁵ Equality

محدودیت مرتبه‌ی موضعی^۱، گزینش‌پذیری^۲ [۱۹ و ۲۵] و پنجره غلظت- صفر^۳: مرتبه‌ی موضعی به ویژگی‌های ریاضی یک ماتریس داده مربوط می‌شود و برای همه‌ی ماتریس‌های داده صرف‌نظر از ماهیت شیمیایی قابل کاربرد است. اگر اطلاعاتی راجع به مرتبه‌ی موضعی گونه‌ها داشته باشیم و آن را در هنگام محاسبه‌ی پروفایل‌های خالص گونه‌ها به‌کار گیریم، محدوده‌ی پاسخ‌های ممکن کاهش خواهد یافت. هنگامی که یک گونه به‌طور گزینشی در داده حضور دارد، پنجره‌های گزینشی آن می‌تواند تحت عنوان محدودیت گزینش‌پذیری اعمال شود. در اغلب موارد اعمال این محدودیت ابهام موجود در پروفایل گونه موردنظر را از بین می‌برد و منجر به حل منحصر به فرد می‌شود. پنجره غلظت- صفر، به مفهوم اینکه در یک محدوده‌ی خاص از غلظت ماتریس داده، غلظت گونه‌ای صفر باشد که می‌تواند به عنوان محدودیت اعمال شود.

۷.۱ تحلیل عاملی^۴

مالینوسکی تحلیل عاملی را به صورت زیر تعریف می‌کند [۲۶]:

"تحلیل عاملی تکنیکی چند متغیره است که با استفاده از فضای عامل‌ها، ابعاد ماتریس داده‌ها را کاهش می‌دهد و با تبدیل آن‌ها منجر به شناخت عامل‌های اصلی می‌گردد."

مراحل تحلیل عاملی به شرح زیر است:

- ۱- تحلیل گونه‌ی اصلی^۵: ماتریس داده‌ها را به بردارهای ویژه^۶ و مقادیر ویژه^۷ تفکیک می‌نماید. این تفکیک می‌تواند بر اساس تکنیک‌های مختلفی از جمله NIPALS^۸ و تجزیه مقادیر منفرد^۹ (SVD) انجام گیرد. در این مرحله ماتریس داده به دو ماتریس امتیاز^{۱۰} و بار^{۱۱} تبدیل می‌شود.
- ۲- مرحله برقراری ارتباط بین عامل‌های خلاصه شده و عامل‌های دارای مفاهیم فیزیکی و شیمیایی است که مهمترین مرحله‌ی تحلیل عاملی به شمار می‌آید. این کار می‌تواند از طریق ضرب کردن فاکتورها در یک ماتریس یا بردار تبدیل و چرخش فاکتورها در فضا انجام شود.

۱.۷.۱ تفکیک مقادیر منفرد (SVD)

¹ Local rank

² Selectivity

³ Zero-concentration window

⁴ Factor analysis

⁵ Principle component analysis

⁶ Eigen vectors

⁷ Eigen values

⁸ Non-Linear Iterative Partial Least Squares (NIPALS)

⁹ Singular Value Decomposition

¹⁰ Score Vectors

¹¹ Loading Vectors

تفکیک مقادیر منفرد [۲۷] یک تکنیک ریاضی برای تفکیک یک ماتریس به بردارهای ویژه به شمار می‌رود. هر ماتریس دو خطی می‌تواند به محصول دو یا چند ماتریس، طبق معادله‌ی زیر تفکیک شود:

$$\mathbf{D} = \mathbf{USV}^T + \mathbf{E} \quad (۱۴-۱)$$

ماتریس‌های \mathbf{U} و \mathbf{S} و \mathbf{V} عامل‌های جزئی‌اند و ویژگی‌های زیر را دارند:
 \mathbf{U} شامل بردارهای ویژه^۱ \mathbf{DD}^T و \mathbf{V} شامل بردارهای ویژه $\mathbf{D}^T\mathbf{D}$ و \mathbf{S} یک ماتریس قطری از مقادیر منفرد^۲ است که جذر مقادیر ویژه \mathbf{DD}^T و $\mathbf{D}^T\mathbf{D}$ است.
 \mathbf{U} و \mathbf{V} ارتونرمال‌اند به این معنی که همگی ستون‌های \mathbf{U} و \mathbf{V} بر هم عمودند و در نتیجه روابط $\mathbf{U}^T\mathbf{U}=\mathbf{I}$ و $\mathbf{V}^T\mathbf{V}=\mathbf{I}$ صدق می‌کنند، \mathbf{I} ماتریس همانی می‌باشد. \mathbf{E} ماتریسی هم بعد ماتریس \mathbf{D} می‌باشد و ماتریس باقیمانده^۳ نامیده می‌شود. ماتریس \mathbf{E} معمولاً حاوی تغییرات تصادفی سیگنال است که در فضای بردارهای ویژه قرار ندارد.

۸.۱ روش لاوتن^۴ و سیلوستر^۵

روش‌های آنالیز نرم نخستین بار توسط لاوتن و سیلوستر در سال ۱۹۷۱ در یک مقاله ۱۶ صفحه‌ای در ژورنال تکنومتریکس معرفی شدند [۹]. کاری که این دو دانشمند ارائه نمودند بیان مسأله تفکیک به زبان ریاضی برای طیف‌های به دست آمده از یک مخلوط دو جزئی ساده بود. بسیار جالب است که چند ایده‌ی اصلی که در این کار معرفی شدند خود سرآغازی برای پیدایش روش‌های مدل‌سازی نرم بودند که پس از آن ارائه شدند. به عنوان مثال ایده‌ی این که، پروفایل‌های گونه‌های خالص ترکیب خطی طیف‌های اندازه‌گیری شده هستند و برعکس، این ایده که فضای ایجاد شده توسط گونه‌های واقعی سازنده‌ی داده با فضای ایجاد شده توسط اجزا اصلی (PCA) یکسان است و ایده‌ی این که یک محدوده جواب داریم و چگونه با اعمال محدودیت‌ها این محدوده را باریک‌تر کنیم، در این مقاله ارائه شدند. الگوریتم روش لاوتن و سیلوستر بر اساس آنالیز اجزای اصلی (PCA) است، مرزهای جواب ممکن با این شرط به دست می‌آیند که همگی مقادیر طیفی و غلظتی گونه‌های خالص باید غیر منفی باشند. با استفاده از این روش تمام جواب‌های ممکن برای هر جز خالص محاسبه می‌شوند، داده‌های موجود در مخلوط دو جزئی روی یک خط قرار می‌گیرند و دامنه جواب

¹ Eigen vector

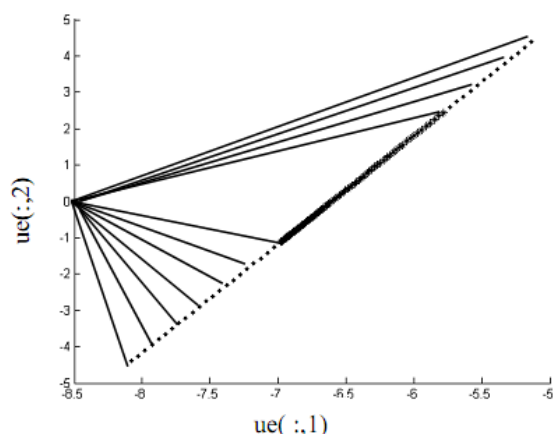
² Singular value

³ Residual Matrix

⁴ Lawton

⁵ Sylvestre

برای هر جز در دو امتداد این خط قرار دارند. به عبارت دیگر برای پیدا کردن پاسخ‌های ممکن برای این داده کافی است در راستای داده‌ها در دو جهت از آن‌ها دور شده و پروفایل‌های طیفی و غلظتی مربوط به این نقاط محاسبه شوند. اگر این پروفایل‌ها هیچ مؤلفه‌ی منفی نداشتند این نقاط می‌توانند داده‌های ما را بازسازی کنند. در مرحله‌ی بعد نقطه‌ی دیگری در همین راستا بررسی می‌شود تا جایی می‌توان این برون‌یابی را ادامه داد که پروفایل‌های مربوط به آن نقاط هیچ مؤلفه‌ی منفی نداشته باشند. جایی که اولین مقدار منفی در پروفایل‌ها مشاهده شد دیگر نمی‌تواند از جواب‌های ممکن برای این سیستم باشد و به مرز بیرونی رسیده‌ایم (آخرین خطوط موربی که تمام پاسخ‌های ممکن را در بر دارند). شماتیک این گفته برای یک داده‌ی سینتیکی در شکل (۳-۱)، آمده است.



شکل ۳-۱ شماتیک روش لاوتن و سیلویتتر. خط پررنگتر مربوط به داده‌های ثبت شده در آزمایشگاه است. تمامی پاسخ‌های ممکن برای این داده با نقطه‌چین‌هایی در دو سمت این خط نشان داده شده است.

چنانچه مختصات یکی از جواب‌های ممکن را در فضای U با T^{-1} نشان دهیم از طریق ضرب آن در ماتریس امتیاز ماتریس غلظتی مربوطه به دست می‌آید. در این صورت T مختصات یکی از جواب‌های ممکن در فضای V^T خواهد بود و با ضرب T در V^T ماتریس طیف به دست می‌آید. این کار از طریق معادلات (۱۵-۱) و (۱۶-۱) انجام می‌شود.

$$C = UST^{-1} \quad (15-1)$$

$$A = TV^T \quad (16-1)$$

توجه داشته باشید که در شکل شماره‌ی (۳-۱) تنها نقاط داده در فضای V نشان داده شده‌اند، چون این داده‌ها از یک سیستم سینتیکی به دست آمده‌اند در این فضا بر روی یک خط قرار گرفته‌اند.