

# Parallel Factor Analysis (PARAFAC)

## آنالیز فاکتورهای موازی

کمومتریکس شاخه‌ای از علم شیمی است که از ریاضیات و آمار به منظور بهینه‌سازی اطلاعات شیمیایی حاصل از اندازه‌گیری‌های انجام شده استفاده می‌کند تا بیشترین اطلاعات شیمیایی را از داده‌ها بدست آورد. از دیدگاه کمومتریکس، داده‌ها اغلب براساس بعد بدست آمده طبقه‌بندی می‌شوند. ساده‌ترین نوع داده، داده‌های مرتبه صفر هستند که فقط یک اسکالر به ما می‌دهد. به عنوان مثال می‌توان به داده‌های جمع‌آوری شده از یک pH متر اشاره کرد. داده‌های مرتبه اول همانند شدت‌های جذب یک محلول در یک گستره از طول موج، تولید یک بردار می‌کند. داده‌های مرتبه دوم نیز مثل داده‌های شدت فلورئورسانس یک محلول، در یک گستره از طول موج نشری و یک گستره از طول موج برانگیختگی تولید یک ماتریس می‌کند. اگر یک بعد دیگر مثل نمونه‌ها را به داده‌های مرتبه دوم اضافه کنیم داده‌های مرتبه سوم به دست می‌آید که تولید یک مکعب از داده‌ها می‌نماید. داده‌های مرتبه سوم با روش‌های THREE-WAY مانند روش PARAFAC تجزیه می‌شوند. PARAFAC یک روش چند بعدی برای تجزیه داده‌هاست که در واقع تعمیمی از PCA دو بعدی به درجات بالاتر می‌باشد. مدل پارافک سه بعدی داده‌ها را به اجزای سه بعدی تجزیه می‌کند یعنی به جای یک بردار اسکور و یک بردار لودینگ - که در PCA داریم - در اینجا یک بردار اسکور و دو بردار لودینگ (و یا سه بردار لودینگ) داریم.

الگوریتم‌های متفاوتی برای تفکیک اطلاعات سه بعدی ارائه شده است. یکی از این الگوریتم‌ها روش کمترین مربعات تناوبی (ALS) می‌باشد که بیش از سایر الگوریتم‌ها در پارافک استفاده می‌شود. اساس ALS قدیمی بوده و به سال ۱۹۳۳ بر می‌گردد که بر اساس تقسیم پارامترها به چندین سری پایه‌گذاری شده است. هر سری از پارامترها با اعمال شرط بر روی پارامترهای باقیمانده به روش کمترین مربعات (LEAST SQUARE) تخمین زده می‌شود. تخمین پارامترها مرتباً تکرار می‌شود تا جایی که تغییری در مقدار پارامترها و یا در فیت مدل بر روی داده‌ها مشاهده نشود. چند روش برای تخمین تعداد اجزای اصلی در پارافک وجود دارد که مهمترین آن‌ها روش تشخیص یکنواختی هسته CORCONDIA می‌باشد. این روش بر پایه مدل TUCKER3 استوار می‌باشد که یک روش مدل سازی دیگری برای آرایه‌های چند بعدی است که در آن علاوه بر ماتریس‌های لودینگ، آرایه دیگری نیز که دارای ابعادی به بزرگی ماتریس  $X$  است محاسبه می‌شود (با ابعاد  $f_i, f_j, f_k$ ). که این ماتریس، ماتریس هسته (مرکزی) نامیده می‌شود. مدل پارافک نسخه محدودتری از مدل TUCKER3 می‌باشد، که در آن ماتریس هسته یک ماتریس دیاگونال با عناصر قطری ۱ و سایر عناصر ۰ می‌باشد.

چنانچه که یکی از ابعاد ماتریس‌های داده‌ها یکسان نباشد یعنی نتوان یک مکعب منتظم تولید کرد در این صورت به جای روش PARAFAC از روش PARAFAC2 استفاده می‌شود. به عنوان یک مورد از این گونه داده‌ها می‌توان به روش HPLC-DAD اشاره کرد. در این روش بعد زمان بازداری به علت تغییر در شرایط ستون یا نمونه‌ها تغییر می‌کند، پس ابتدا باید آن را به یک مکعب منتظم تبدیل کرد و سپس طبق روش PARAFAC داده‌ها را تجزیه کرد.