一种分析混合物成分的方法

技术领域

本发明涉及化学分析领域，具体地，涉及一种分析混合物成分的方法。

背景技术

现有分析方法中，分析混合物中的各成分主要是利用各种色谱分析方法，即利用混合物中各成分在各种色谱中流动相和固定相上的化学、物理性质的不同，使得各成分在色谱中移动速度产生差别，从而使各成分在不同的时间经过检测器（如紫外、红外、质谱等），获得各种谱图，从而分离、分辨混合物中的各成分。

混合物全成分的分析状：

对复杂混合物中各成分进行分析，常用分析方法是在多次不同条件（例如改变温度、改变温度变化方式、改变流速、使用不同的固定相（如更换色谱柱）和不同的流动相等）的色谱分析中，每次将一部分的混合物分离，然后综合各个不同条件下的结果，得到较为完全的分析结果。现有方法经过多次试验，仍然不能完全知道是否全部分离了成分。

现有分析方法对复杂混合物的全成分分离非常耗时、耗力和耗费，而且不通用。通常对一种复杂混合物的分析，例如大蒜中成分的分析，需要多年的研究，同时需要几百万甚至几千万的投入，并需要很多专业人员。而且，由于不同混合物中的成分不同（matrix），所以现有分析方法也不是通用的。例如大蒜成分的分析方法，对于洋葱就不适合了，需要重复分析，还是需要大量资金、时间和人员的投入。

现有分析方法对复杂混合物全成分分析的最大困难为：分离，就是如何将混合物中各个成分一个一个分离开。对于稍微复杂的混合物，将混合物在几次分析中全部分离开是不可能的，出现重叠谱图（混合物谱图）是不可避免的，更不要说有着成千上万成分的中药和天然产物了。如何尽快地分离开混合物中的各个成分，从而对其中各个成分进行分析，这是个世界性的难题。

标的物的分析现状：

对于含有标的物的分析，例如食品、药物或环境安全等的分析，现有分析方法一般需要将标的样品尽量提纯，或者将样品中的标的物做物化处理，例如氘代反应等；这些前处理一般都比较繁琐、费事。然后，处理好的样品需要经过色谱分离，将标的物从背景化合物中（martrix）分离开，然后才能定性和定量。

由于各种标的物不同，导致检测、分析各种标的物的过程中，所需要的仪器和条件等都不一样，所以一般的分析中心只能对于一些特定的标的物做分析检测，而且分析成本都会比较高。

标的物的分析，其根本的问题，和混合物全成分分析的根本问题一样，都是分离的问题。

所以，解决了分离的问题，就是解决分析混合物成分的核心问题。

熵最小算法（Entropy Minimization Algorithms）

在分析仪器和各种分析方法中，大量应用了各种数学方法，帮助解决分析中的各种问题，例如基线的问题等。这些数学方法统称为化学计量法。

有一类化学计量法主要是用于从混合物谱图中提取出纯成分的信息（例如SIMPLSMA，IPCA，OPA-ALS）。但是这些方法需要一些先决条件和信息，例如重叠谱图中每个纯成分的峰型是对称的（例如高斯对称），每个纯成分峰之间需要有一定的间隔，重叠程度也不能太大等；另外还需猜测混合物中有几个成分等。由于先决条件太多，这些数学方法对混合物中的纯成分分离帮助有限，只能起辅助作用。尤其是对于复杂混合物，例如中药和天然产物的样品，里面的成分信息是猜不出的，分析产生的重叠谱图也不能达到这些方法的应用要求。

熵最小算法（EM）是基于Shannon Entropy（香农熵）发展而来的。在没有提供任何成分信息的情况下，只需要混合谱数据，就能从混合谱中提取出每个纯谱。Shannon Entropy原理最早发表于1948年[C.E. Shannon, The Bell System Technical Journal, 27 (1948) 379-423.]，它是一个信息领域的学术用语，用于衡量随机参数的不确定性。

最早将Shannon Entropy应用于化学分析的是Marc Garland [Y.Z. Zeng, M. Garland, Analytica Chimica Acta, 359 (1998) 303-310.]，他在2002年发表了BTEM (Band-Target Entropy Minimization) [W. Chew, E. Widjaja, M. Garland, Organometallics, 21 (2002) 1982-1990.]，即锁峰熵最小方法。该方法是用于分离简单的红外混合谱。BTEM需要人为指导操作，主要应用于研究领域。

2003年张华俊等发表了tBTEM方法（Weighted Two-Band Target Entropy Minimization [H.J. Zhang, M. Garland, Y.Z. Zeng, P. Wu, J Am Soc Mass Spectrom, 14 (2003) 1295-1305.]），即加权多重锁峰熵最小方法，该方法开始主要应用在质谱上，同时一种自动分析、去除重合谱的方法也一同发表了。

2006年张华俊等发表了MREM [H.J. Zhang, W. Chew, M. Garland, Applied Spectroscopy, 61 (2007) 1366-1372.]（Multi-Reconstruction Entropy Minimization）方法，即多重建熵最小方法。该方法用局部优化方法代替了全局优化方法，而且不需人为指定搜索范围，真正实现了自动搜索纯谱的功能。

2009年熵最小方法顺利应用于紫外谱图解析 [F. Gao, H.J. Zhang, L.F. Guo, M. Garland, Chemometrics and Intelligent Laboratory Systems, 95 (2009) 94-100.]。由于紫外谱图非常宽，峰形不对称，各成分重叠非常厉害，即使2~3个成分的紫外混合物谱图也是非常难以解析的。

将熵最小系列方法（BTEM、tBTEM和MREM等）应用于混合谱中，能在不知道任何成分的情况下，仅利用混合物的谱图信息，就能从混合谱中重建出纯谱。它们已经成功应用于红外、紫外、质谱和核磁共振图谱等谱图中。

EM系列方法应用的前提是：1）混合谱谱图的数目应大于混合谱中成分的数目；例如，一个混合物中含有10个成分，那么经计算机采样得到的混合谱谱图的数目就应该大于10；2）

各成分在每个混合谱谱图中的浓度比例，应该是不同的；如果混合物谱图中的任意两个或者多个成分的比例在不同的采样时间都保持一致，那么这几个成分对于熵最小算法来说，就是不可分辨的。现有的文献中，各种EM系列方法的应用，是利用检测器检测一个动态反应的不同过程，在不同的反应时间采样，从而得到很多不同的混合谱，从而使得各个混合谱中各成分的比例不一致。例如：在一个不平衡的化学反应过程中，于20个不同的时间采样，得到20个混合谱；由于反应随着时间推进，所以反应物和生成物的浓度随着时间发生变化，从而导致反应物和生成物的在各个采样点的浓度比例不一致。

然而实际的化学分析中，一个混合物样品，例如一个中药样品或者一个牛奶样品，里面各个成分的浓度比例是恒定的，所以就无法满足上述EM系列方法应用的两个条件。如何从一个恒定浓度比例的样品中，得到很多不同比例的混合物谱图，从而利用EM方法进行分析，这又是一个亟待解决的问题。

发明内容

为了解决上述技术问题，本发明的目的是提供一种分析混合物成分的方法，该方法将色谱分离技术和熵最小系列算法结合起来，可以快速分析混合物的各成分。

本发明提供的分析混合物成分的方法，包括：

（1）将混合物样品，利用色谱技术进行分离，得到初步的色谱图；

（2）对步骤（1）得到的初步的色谱图中的任意一个区间，根据需要进行采样，得到一系列在不同保留时间的谱图，称之为混合物谱图（即混合谱）；

（3）利用熵最小算法的系列方法（即EM系列方法），从步骤（2）得到的混合物谱图中，计算得到各重建纯谱及其对应的伪浓度。

其中，步骤（1）中，得到的初步的色谱图中，包括纯成分（已经分离出的成分）的峰和混合物的重叠峰。

进一步地，对于纯成分的峰，利用现有常规方法已经能够进行分析与鉴定，一般不需要利用本发明方法后续的步骤进行分析和处理，当然使用者也可以利用本发明方法后续的步骤进行分析；对于混合物的重叠峰，需要利用本发明上述方法后续的步骤（上述步骤（2）和步骤（3）进行分析。

其中，步骤（1）得到的初步的色谱图中的任意一个区间，是指初步的色谱图的一段，或者全部。

进一步地，步骤（2）中的数据采样，指将需要的区间数据从仪器产生的数据中提取。采样是否均匀由仪器本身决定。

其中，步骤（1）中，色谱技术为：气相色谱技术、液相色谱技术或者两者的组合。

其中，步骤（1）中，利用色谱技术进行分离，经检测器得到混合物的谱图。

进一步地，所述检测器选自：质谱、红外、紫外、荧光、分光光度仪和核磁共振检测器中的一种或几种的组合。

其中，步骤（3）中，熵最小算法的系列方法是个总称，其包括基于Shannon Entropy（香

农熵）的各种方法，包括但不限制于BTEM(Band-Target Entropy minimization)、tBTEM(Weighted Two-Band Target Entropy Minimization)或MREM(Multi-Reconstruction Entropy Minimization))等方法中的一种或几种的组合。

进一步地，利用BTEM或者tBTEM方法，每次运算，可以得到一个重建纯谱；利用MREM方法，每次运算，可以得到多个重建纯谱。

本发明中的"重建纯谱"指的是：利用熵最小算法的系列方法（即各种熵最小算法），从混合物谱图中，经过计算得到的纯谱。

本发明中的"各重建纯谱对应的伪浓度"指的是：利用熵最小算法的系列方法，在得到重建纯谱的时候，同时会计算得到一个和该重建纯谱相关的浓度（即叫做"伪浓度"）。

其中，步骤（3）中，对于含有未知物的混合物样品的分析，需要尽可能知道步骤（2）得到的混合物谱图中的成分信息的时候，利用熵最小算法的系列方法，在不同的参数下，反复多次计算后，得到尽可能多的重建纯谱。

上述方法中：

不同的参数：指锁定不同的峰，使用不同的目标函数和罚函数或使用不同的优化方法等。如：在一个 m/z 从 100-200 的混合物谱图中（间隔是 1 m/z），该数据共有 101 个 m/z 通道；可以将每个通道锁定，然后计算，或者任意两个通道进行计算；或者只对通道值大于 30%最大峰的通道进行计算等。

计算多次：通过不同的参数设定，进行计算，每次都得到一个重建纯谱（BTEM 或者 tBTEM 方法）或者多个重建纯谱（MREM 方法）。由于不同的参数设定，导致多次的计算。

尽可能多的重建纯谱：由于多次计算后，会有很多重建纯谱，有些纯谱是相同的，所以可以利用算法去除，最后每个不同的纯谱都只有一个。

其中，步骤（3）中，对于含有标的物/已知物的混合物样品的分析，基于标的物/已知物的标准谱图，利用熵最小算法的系列方法计算，构建使用不同参数的重建方法，得到该标的物/已知物的重建纯谱；在多次的重建中，如果重建纯谱和标的物/已知物的标准谱图相比较是一致的，而且该重建纯谱的伪浓度是有意义的，则该混合物中含有该标的物/已知物，反之则不含有。

进一步地，多次的重建，例如：根据标的物纯谱，对于某个数据通道的值>30%最高值的，就进行锁住，然后做计算；或者可以每个通道都锁住一次，然后进行计算。

进一步地，重建纯谱的伪浓度是有意义的，是根据经验，在一般的色谱柱分析中，一般峰型是较对称的单峰，而不是如同随机值一样的。

本发明所述的方法，在步骤（3）后，可以将得到的重建纯谱和标准数据库中的谱图进行比对，确认它们的信息；若比对后重建纯谱和标准数据库中的某谱图一致，则说明重建纯谱代表的是标准数据库中的某谱图所代表的成分。若比对后重建纯谱和标准数据库中的谱图都不一致，则表明发现了标准数据库中没有记录的化学成分；后续可以再寻找其他数据库进行比对，或根据需要利用后续的各种分析方法鉴定该成分。

本发明提供的方法，有益效果如下：

1）主要克服了现有分析方法对成分分离的依赖，使得分离不再重要，克服了现有分析方法的技术瓶颈；

2）将色谱技术和熵最小算法结合，克服了彼此的缺点；

3）将色谱技术和熵最小算法结合连用后，不需要彻底地分离混合物中各个成分，同时也达到了熵最小算法使用的先决条件；

4）扩大了熵最小算法的应用范围，可以分析各成分比例恒定的混合物样品；

5）可以分析混合物样品中的未知成分，也可以分析含有已知成分/标的物的混合物，进而可以对各成分进行定性与定量；

6）是一个通用的分析方法，不同的混合物样品均适用；

7）速度快，因为不需要花非常多的时间在分离上；成本低，不需要十分优良的分离设备，对人员要求也低。

总之，本发明提供的分析混合物成分的方法，将色谱技术和熵最小算法结合，主要克服了现有分析方法对成分分离的依赖，使得分离不再重要；不需要彻底地分离混合物中各个成分，同时也达到了熵最小算法使用的先决条件；并且方法通用、速度快、高效、成本低，对人员要求不高，具有巨大的应用前景。

附图说明

图1为EM系列方法应用于混合物样品（全部都是未知物）的分析流程图。

图2为EM系列方法应用于含有标的物/已知物的混合物样品的分析流程图。

图3为实施例1中，航空煤油样品的TIC谱图，出峰时间为14~15分钟。

图4为实施例1中，航空煤油样品，从出峰14~15分钟的TIC谱图，总共采样了176次数据的混合谱，集中显示的3-D质谱图。

图5-1为实施例1中，线型正十一烷烃的EI-MS纯谱谱图（从日本AIST数据库获得）。

图5-2为实施例1中，线型正十二烷烃的EI-MS纯谱谱图（从日本AIST数据库获得）。

图6-1为实施例1中，通过锁定峰m/z=57和170，运用EM方法计算得出的纯谱。

图6-2为实施例1中，通过锁定峰m/z=57和170，运用EM方法计算得出的纯谱对应的伪浓度。

图7-1为实施例1中，通过锁定m/z=128，运用EM方法计算得出的纯谱。

图7-2为实施例1中，通过锁定m/z=128，运用EM方法计算得出的纯谱对应的伪浓度。

图8为实施例2中，挥发物样品的TIC谱图，m/z=800-850秒。

图9为实施例2中，挥发性物质样品的3-D质谱图，TIC = 826.22 - 831.707秒。

图10-1为实施例2中，当锁住m/z = 91的峰，用EM计算出的重建纯谱。

图10-2为实施例2中，当锁住m/z = 91的峰，用EM计算出的重建纯谱对应的伪浓度。

图11为实施例2中，锁住m/z=91的峰后，重建纯谱和NIST数据库中的标准谱图相匹配的结果；其中丰度正值部分为重建纯谱，丰度负值部分为NIST数据库中标准谱图。

图12-1为实施例2中，当锁住m/z = 71的峰，用EM计算出的重建纯谱。

图12-2为实施例2中，当锁住m/z = 71的峰，用EM计算出的重建纯谱对应的伪浓度。

图13为实施例2中，锁住m/z = 71的峰后，重建纯谱和NIST数据库中的标准谱图相匹配的结果；其中丰度正值部分为重建纯谱，丰度负值部分为NIST数据库中标准谱图。图14-1为实施例2中，当锁住m/z = 105的峰，用EM计算出的重建纯谱。

图14-2为实施例2中，当锁住m/z = 105的峰，用EM计算出的重建纯谱对应的伪浓度。

图15为实施例2中，锁住m/z = 105的峰后，重建纯谱和NIST数据库中的标准谱图相匹配的结果；其中丰度正值部分为重建纯谱，丰度负值部分为NIST数据库中标准谱图。

图16-1为实施例2中，当锁住m/z = 57和85的峰，用EM计算出的重建纯谱。

图16-2为实施例2中，当锁住m/z = 57和85的峰，用EM计算出的重建纯谱对应的伪浓度。

图17为实施例2中，锁住m/z = 57和85的峰后，重建纯谱和NIST数据库中的标准谱图相匹配的结果；其中丰度正值部分为重建纯谱，丰度负值部分为NIST数据库中标准谱图。

图18为实施例2中，经过矫正的4个重建纯谱的浓度，和重建总TIC浓度，以及、实际TIC浓度的汇总比较图。

图19为值空间和优化方法示例：在一次优化中，BTEM/tBTEM熵最小算法只会找到最下面的一个纯谱；而MREM算法就会找到全部的多个谱图。

具体实施方式

以下实施例用于说明本发明，但不用来限制本发明的范围。

本发明所用的物料，均为可以从市场上购买得到的常规物料。本发明中未提及的其他操作步骤，均为本领域的常规操作。

以气相色谱-质谱连用方法（GC-MS）为例（流程图如图1所示，含有未知物的混合物样品的分析）:

将一个混合物样品注入气相色谱（GC）仪，混合物中含有A、B、C……和Z等成分，由于各成分自身物理化学性质（极性、分子大小等）的不同，它们在色谱柱中移动的速度不一样，从而导致各个成分到达检测器（质谱）的时间先后不同，而出现不同的峰。图1中，峰1、2和3是纯成分（已经分离出的成分）的峰；而峰4是一个混合物的重叠峰，里面的成分没有被分离开。现有的分析方法，需将一个成分和其他成分分离开，然后才能定性（例如峰1、2和3）。对于没有分离开的成分（例如重叠峰4），现有常规方法就需要将同样的样品重新在不同条件下做分离，经过反复非常多的次数，才能将各个成分在不同的分离过程中分离开。对于不同的样品，由于成分不同，所以分析方法不能通用。现有的分析方法，大部分的时间和精力都花在如何分离上，而且需要复杂的样品前处理过程，以减少样品中的成分数，从而降低色谱分离难度。

本发明的方法中，先用色谱技术（GC）分离混合物样品中的成分，经检测器（质谱）得到混合物的谱图（即初步的色谱图）。对于纯成分的峰（已经分离出的成分的峰，例如峰1、2和3），不再做任何处理。而对于重叠峰4这样的混合物峰，本发明就利用数据采样，即在不

同的采样点进行数据采样(采样不一定是均匀的),从而得到一系列的混合谱(即混合物谱图);然后利用EM方法对混合谱进行多次操作,从而得到不同的重建纯谱及其对应的伪浓度。

将一个成分浓度比例固定的样品,利用色谱的分离能力,经过色谱技术分离后,在检测器端,就能得到一系列的含有不同成分、不同浓度比例的纯成分或者混合物的谱图。对于未分离好的混合成分的重叠峰,经过计算机多次采样(在不同时间点进行数据采样),得到的混合物谱图数目就会大于重叠峰所代表的混合物中含有的成分数。不同的化合物在色谱柱中的移动速度,不论色谱柱的分离能力高低,总有差异,所以在任何一个采样点,各个成分之间的比例是不一样的,这样就满足了EM系列方法应用的前提。

本发明的方法中,分离混合物的各成分主要是依赖EM系列方法重建出各成分的纯谱,所以对色谱技术中色谱柱的分离能力要求不高,但是好的色谱分离是欢迎的,因为这样能够减少混合物重叠峰,从而减少需要EM系列方法数学处理的量。对于不同的混合物样品,它们中的成分千差万别,同一种色谱分离方法将导致不同的分离结果,这是现有分离技术的问题,也是现有分离技术对不同样品不通用的原因;但是本发明的技术由于只是对混合物重叠峰操作,所以不同混合物样品的不同分离结果,对本发明的方法没有进行区分;也就是说,本发明的方法只要对不同的混合物重叠峰处理就行,而不在乎它们出现在什么地方。所以本发明所提供的方法,是一个通用的分析方法,对样品的成分组成情况不敏感;同时本发明的方法也可以大大降低样品前处理的过程;因为未经过前处理的样品,经色谱分离,会多一些混合物重叠峰,只是增加计算量,本发明方法对此没有任何敏感性。

利用本发明的技术,从混合谱(混合物谱图)中能够重建得到纯谱,本发明的方法并不能知道这些纯谱代表的是什么物质。在有可能的情况下,例如在运用气相色谱-电子电离质谱方法中(GC-EI-MS),由于EI-MS有庞大的成分数据库(例如NIST质谱数据库),所以将本发明方法计算得到的纯谱谱图和该成分数据库以及其他的成分数据库进行比对,就能容易地知道的重建出的纯谱代表的物质是什么了。

利用本发明的技术,对于混合物样品中的标的物/已知物分析(即含有标的物/已知物的混合物样品的分析)也非常有用,只需要利用标的物/已知物的标准谱图,设计EM系列方法的应用方法(例如锁住标的物/已知物的特征峰或特征谱图区间等),在没有分离好的谱图上,就可以计算出是否含有标的物/已知物和该标的物/已知物的伪浓度;在多次的重建中,如果在某一次结果中,重建纯谱和标的物/已知物的标准谱图是一致的,而且该重建纯谱的伪浓度是有意义的,则可以判断该混合物中含有该标的物,反之则不含有。这样可以大大降低对分离的要求,加快速度;该应用对于质检、检测和监测行业都非常有用。以气相色谱-质谱连用方法(GC-MS)为例,流程图如图2所示的为含有标的物/已知物的混合物样品的分析。

本发明的方法,将EM系列方法应用在标的物/已知物分析中,大大降低了对分离的要求,由于不需要将标的物/已知物和其他成分分开,所以对于不同的混合物样品,也就变得不敏感了;即不再需要设定各种不同的分析程序,所以本发明的方法是一个快速、通用的方法。

一、本发明中所使用的熵最小算法(BTEM,tBTEM和MREM)解释如下:

当一个样品，经过GC-MS后，经计算机采样得到一套数据 $A_{k\times v}$，其中 k 是计算机采样次数，例如从出峰时间10-12分钟之间，每隔0.1秒采样，共采样21次。v 是数据通道数，例如被采样质谱的质量电荷比（m/z）是10-100，间隔是1个m/z，共91个数据通道。

下面，将算法的标准程序按照先后次序列出：

1）将数据集 $A_{k\times v}$ 按照方程（1）进行奇异值分解（SVD），然后将右奇异矩阵 $V^T_{v\times v}$ 没有物理意义的部分除去，将对角矩阵 $S_{k\times v}$ 中的全零部分去除，得到 $V^T_{k\times v}$ 和 $S_{k\times k}$，矩阵 U 没有用到。

$$A_{k\times v} = U_{k\times k} \times S_{k\times v} \times V^T_{v\times v} \qquad (1)$$

2）对矩阵 $V^T_{k\times v}$ 中的每个行向量进行观察，那些看起来像背景噪声的行向量可以去除，例如从第 j 行向量以后看起来都是噪声，我们就把 $(j+1)$ -k 行向量都去掉，得到 $V^T_{j\times v}$。

3）从那些保留下来的 $V^T_{j\times v}$ 中，通过检视这些 j 个行向量，使用者可能对某一个或几个峰，或者是某一区间有兴趣；例如峰 m/z=91，或者 m/z=90-100 感兴趣；使用者就可以将这些峰或区间锁定（target），并开始运算。

4）首先，一个随机向量 $T_{1\times j}$ 由计算机生成，该 T 向量将由所使用的优化算法自动更新，直到找到最优值。通过 T，一个和纯谱有关的向量 a^{est} 通过方程（2）得到。a^{est} 将根据 T 的变化而变化，直到优化结束方程；最后得到的 a^{est} 我们认为就是纯谱。方程（2）中的 S 矩阵（见步骤1）可用，也可不用，根据实际需要而定。

$$\mathbf{a}^{est}{}_{1\times v} = \mathbf{T}_{1\times j} \times (\mathbf{S}_{j\times j} \times \mathbf{V}^T_{j\times v}) \qquad (2)$$

5）将 a^{est} 用锁定的峰或区间中的最高峰，按照方程（3）单位化。单位化后的向量，我们叫 $\hat{a}$。这个过程就是我们的“锁峰”过程。其中，a′ 和 a″是锁定的峰和区间。

$$\hat{a}_{1\times v} = \frac{a^{est}_{1\times v}}{\max(a') + \max(a'')} \qquad (3)$$

6）利用的得到的 a_{hat} 构建优化方法的目标函数和罚函数，方程（4）是总的目标和法函数，具体目标函数和罚函数，根据不同类型的谱图，有不同的表达方式，详细请参考如下相关文献（文献①、②和③）。其中 P 函数是罚函数。通过方程（5），计算得到伪浓度信息。

$$min(G) = \sum_v \hat{a}_{1\times v} + P(\hat{a}_{1\times v}, \hat{c}_{k\times 1}) \qquad (4)$$

$$\hat{c}_{k\times 1} = A_{k\times v} \times \hat{a}^T_{v\times 1} \times (\hat{a}_{v\times 1} \times \hat{a}^T_{v\times 1})^{-1} \qquad (5)$$

7）检测目标函数的值，看是否已经达到优化终点；如果达到了，那么就输出纯谱 a^{est}，和伪浓度 C_{kX1}。如果没有达到优化终点，那么就将由优化方法，重新生成一个 T_{1Xj}，返回到步骤4）到7）重新计算，直到优化结束。

重复步骤3）到7），就可以锁定不同的峰或区间，运行算法，得到不同的纯谱和相关浓

度，然后就可得到混合谱中的其他，或者是全部纯谱信息。

上述文献为：

①Chew, W., E. Widjaja, and M. Garland, Band-target entropy minimization (BTEM): An advanced method for recovering unknown pure component spectra, application to the FTIR spectra of unstable organometallic mixtures. Organometallics, 2002. 21(9): p. 1982-1990.

②Zhang, H.J., et al., Weighted two-Band Target Entropy Minimization for the reconstruction of pure component mass spectra: Simulation studies and the application to real systems. Journal of the American Society for Mass Spectrometry, 2003. 14(11): p. 1295-1305.

③Zhang, H., W. Chew, and M. Garland, The multi-reconstruction entropy minimization method: Unsupervised spectral reconstruction of pure components from mixture spectra, without the use of a Priori information. Applied Spectroscopy, 2007. 61(12): p. 1366-1372.

二、优化方法的使用：

熵最小系列算法（BTEM，tBTEM 和 MREM），都需要用到优化方法（得到 T）。

首先，熵最小方法利用 SVD 后的 V^T 数据，利用优化目标方程（4），构建一个多维值空间；然后利用优化方法，在这个值空间里搜索纯谱。对于锁住的不同峰或区间，其构建的多维值空间不同。

如图 19 所示，在搜索方式上，BTEM/tBTEM 认为全局最小点是纯谱的位置，所以用的是全局优化方法，每次优化，只能得到一个纯谱。后来张华俊发现（上述文献③），在值空间中的各个局部最小点，都对应一个纯谱，所以研发了 MREM；MREM 使用局部优化方法，每次优化，在每个局部最小点，都有一个纯谱；所以一次优化，能够得到多个纯谱，所以速度大大加快，能力大大加强。将 BTEM，tBTEM，MREM 方法混合使用，效果更强；能够找到100%被其他谱图淹没的谱图，而且能够应用于紫外这样的高难度体系。

实施例1：在气质联用分析仪上应用本发明的方法分析航空煤油

一、实验条件

一个航空煤油样品（市场上购得），在安捷伦的气相色谱质谱联用分析仪上分析，色谱柱型号是HP5-MS，温度控制是先30℃恒温5分钟，然后从30℃升温到300℃，升温速度是10℃/分钟。由于航空煤油是个复杂的混合物，里面主要是很多结构类似的烷烃和芳香烃成分，还有很多成分属于同分异构体，所以混合程度非常复杂；即该航空煤油体系中各个成分用气相色谱质谱联用分析仪无法完全分离开。

该混合物经过上述仪器分析，得到TIC（Total Ion Current）谱图，其中从出峰14~15分钟的TIC谱图见图3。对这个时间段的谱图进行数据采样，总共采样了176次数据的混合谱，得到的一系列混合物谱图见图4（显示在一张3-D质谱图中）。其中质谱数据点是从50~200 m/z，间隔为1 m/z（每次采样时间是指一个特定的TIC时间，但是每次采样后的质谱，数据点就是50~200 m/z，间隔为1 m/z）。

二、快速发现航空煤油混合物中的已知成分

航空煤油样品中，含有很多烷烃。从日本的AIST（National Institute of Advanced Industrial Science and Technology, Japan.）有机物数据库中找到了正十一烷烃的EI-MS标准谱图（见图5-1）。其中正十一烷的最高峰在m/z=57处，分子量峰在156 m/z处。另外一个烷烃，正十二烷的最高峰也是在m/z=57处，而分子量峰在170 m/z处。

对于采样好的混合物数据，应用EM算法，希望了解在TIC=14~15min的区间是否有正十一烷。EM算法将峰m/z=57单独锁定（具体用BTEM方法），或者将峰m/z=57和156一起锁定（具体用tBTEM方法），每种锁定方法都经过5次反复运算。结果发现，应用两种算法，每次计算出的重建纯谱，与该成分的标准谱图进行比对，没有找到相一致的图谱。所以认为，在TIC=14~15分钟之间，没有正十一烷烃。这个结果和实际实验结果吻合，因为用纯的正十一烷，在相同实验条件下做标定，该成分实际出峰时间在12.75分钟左右。

对于 TIC=14~15min 的区间，是否含有正十二烷，也利用上述 EM 方法试图找出。将峰 m/z=57 单独锁定，或者将峰 m/z=57 和 170 一起锁定，每种锁定方法都经过 5 次反复运算；每次计算出的重建纯谱和正十二烷的标准谱图（图 5-2）相符。

其中锁定峰 m/z=57 和 170 的计算出的纯谱见图 6-1，计算出的伪浓度峰显示（见图 6-2），正十二烷烃出峰时间在 14.37 分钟左右。从上述图 3 实验结果中可以看出，TIC=14.33~14.43 分钟之间，有一个很明显的混合物的重叠峰，这个重叠峰至少包含有两个成分，而且该峰重合度太大了，已经分辨不出其中的纯谱。而根据上述 EM 的计算结果，运用本发明方法则顺利的将正十二烷的纯谱重建出来，而且同时得到该正十二烷的伪浓度。用纯的正十二烷在相同的实验条件下做标定，其出峰时间就在 14.38 分钟左右。和 EM 计算的结果吻合。

通过该实施例，若想知道混合物样品中是否含有已知成分或标的物，对其进行分析，只需要根据已知物/标的物的标准谱图，在没有分离好的混合物谱图上，通过上述方法就能很快得出是否有该已知物/标的物；如果有的话，可以计算得出它们的浓度（需要根据浓度标准曲线来标定）和出峰时间是多少。

因此，本发明可用于快速分析，对于质检和质监方面非常有用；样品不需要分离好，不同的样品，分析方法类似；也不需要复杂的样品前处理，所以本发明效率非常高。

三、快速发现混合物中的未知成分

经过查看实验得出的 TIC 峰（图 3），发现在 TIC=14.25min 左右有个峰，但是该峰和左右的几个峰重叠了，无法得到该峰所代表的纯成分的峰。在该 TIC=14.25min 的地方，采样得到的质谱混合谱上（采样方法同上），观察到有个 m/z=128 的峰。利用 EM 方法，锁定峰 m/z=128，重建纯谱和其伪浓度如图 7-1 和图 7-2 所示。

但是，如果没有和数据库中纯成分的标准谱图进行比对，无法知道该重建纯谱代表什么化合物。当将重建纯谱和 NIST 标准数据库中的图谱匹配，确定了该重建纯谱和芳香烃萘的标准质谱图谱一样，从而确定在 TIC=14.25min 的峰是化合物萘的峰。在同样的条件下，用纯萘做标定后，萘的 TIC 出峰时间和本实施例的伪浓度峰出峰时间相吻合。

本实施例中上述快速发现混合物中的未知成分的实验说明，对于混合物样品中的未知物，并不需要很好的色谱分离，就能在未知是什么成分的情况下，仅仅根据混合谱中给出的信息（例如质谱峰），就可以快速找出未知成分；进一步若通过和标准谱图数据库进行比对，就能知道各成分的结构信息了。

实施例2：在气质联用分析仪上应用本发明的方法分析挥发性物质

一、实验条件：

一个海藻油经加氢催化得到的航空煤油样品（通过常规方法实验得到），在安捷伦的气相色谱质谱联用分析仪上分析，色谱柱型号是HP5-MS，温度控制是先30℃恒温5分钟，然后从30℃升温到300℃，升温速度是10℃/分钟。

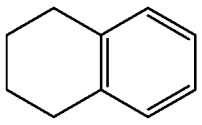
该混合物经过上述仪器分析，得到TIC（Total Ion Current）谱图，其中从出峰800~850秒的TIC谱图见图8。图8中可见，在827~832秒的区间有峰，但是峰型不对称，不知这个峰是否是纯成分的峰。于是对827~832秒的谱图进行数据采样，总共采样了17次数据，得到的一系列混合物谱图见图9（显示在一张3-D质谱图中）。其中质谱数据点是从50~150 m/z，间隔为0.5 m/z。

二、熵最小算法应用：发现未知成分并定性

锁峰 m/z = 91:

利用熵最小算法（具体为 BTEM 算法），将峰 m/z = 91 锁住后进行计算，得到的重建纯谱和其对应的伪浓度如图 10-1 和图 10-2 所示。可以看出，重建纯谱的伪浓度，有比较好的对称性。

重建纯谱的数据，用 MREM 软件导出成*.msp 格式后，导入 Agilent（安捷伦）的气相质谱，和仪器自带的 NIST 数据库中的标准谱图比对后，结果如图 11 所示。匹配结果显示重建纯谱代表的是化合物为 $C_{10}H_{12}$，结构式为：

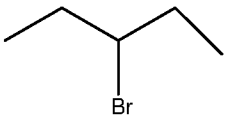


。

锁峰 m/z = 71:

利用熵最小算法（具体为 BTEM 算法），将峰 m/z = 71 锁住后进行计算，得到的重建纯谱和其对应的伪浓度如图 12-1 和图 12-2 所示。可以看出，重建纯谱的伪浓度，有不错的对称性。

重建纯谱的数据，用 MREM 软件导出成*.msp 格式后，导入 Agilent 的气相质谱，和仪器自带的 NIST 数据中的标准谱图比对后，结果如图 13 所示。匹配结果显示重建纯谱代表的化合物是 $C_5H_{11}Br$，结构式为：

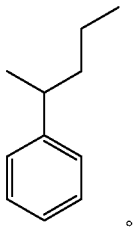


。

锁峰 m/z = 105:

利用熵最小算法（具体为 BTEM 方法），将峰 m/z = 105 锁住后进行计算，得到的重建纯谱和其对应的伪浓度如图 14-1 和图 14-2 所示。可以看出，重建纯谱的伪浓度分布有点奇怪（不太对称），但是也是有可能的。

重建纯谱的数据，用 MREM 软件导出成*.msp 格式后，导入 Agilent 的气相质谱，和仪器自带的 NIST 数据中的标准谱图比对后，结果如图 15 所示。匹配结果显示重建纯谱代表的化合物是 $C_{11}H_{16}$，结构式为:

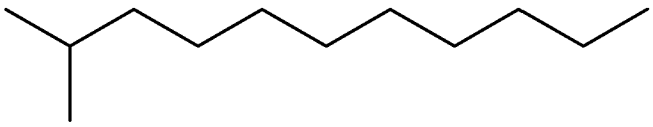


。

锁峰 m/z = 57 和 85:

利用熵最小算法（具体为 tBTEM 算法），将峰 m/z = 57 和 85 锁住后进行计算，得到的重建纯谱和对应的伪浓度如图 16-1 和图 16-2 所示。

重建纯谱的数据，用 MREM 软件导出成*.msp 格式后，导入 Agilent 的气相质谱，和仪器自带的 NIST 数据中的标准谱图比对后，结果如图 17 所示。匹配结果显示重建纯谱代表的化合物是 $C_{11}H_{26}$，结构式为:



。

三、定量分析:

将上述四个重建纯谱的伪浓度经过矫正后，然后将它们的重建 TIC 浓度和实验得出的实际的 TIC 浓度比较（如图 18 所示）。结果显示，重建总 TIC 浓度和实际的 TIC 浓度所代表的峰的形状比较一致。它们的面积数值如下表:

重建纯谱和原始谱图的 TIC 浓度面积的关系		
原始谱图的 TIC 浓度面积 :		10835841
重建纯谱总 TIC 浓度面积 :		10668662
重建总面积 / 原始面积 =		98.5%
各重建纯谱的 TIC 浓度面积信息		
No	TIC 浓度面积	占全部重建浓度的百分比

线 1	2484060	23.3 %
线 2	4142366	38.8 %
线 3	1388395	13 %
线 4	2653840	24.9 %

在图 18 中，各曲线的含义分别表示为：粗实线代表实际（或原始）TIC 浓度；粗虚线表示 4 个重建浓度的简单加和，用于和实际 TIC 浓度做比较；线 1 代表锁峰 m/z=57 和 85 后经过矫正的重建纯谱的 TIC 浓度；线 2 代表锁峰 m/z=91 后经过矫正的重建纯谱的浓度；线 3 代表锁峰 m/z=71 后经过矫正的重建纯谱的浓度；线 4 代表锁峰 m/z=105 后经过矫正的重建纯谱的浓度。

四、结果讨论：

利用化学分析方法，发现混合物中的未知成分，一直是比较困难而繁琐的事情。本实施例说明，只利用混合物谱图中的质谱峰（例如 m/z = 105），然后利用熵最小算法，就能方便地发现未知成分，得到未知成分的重建纯谱；然后利用数据库中的标准谱图进行比对，就能知道这些未知成分的纯谱所代表的物质。本实施例也显示，利用熵最小方法，能够很好的重建各成分的 TIC 浓度；因此就能够用常用的气相质谱方法定量各个成分的浓度。

实施例 3：MREM 算法在气质联用分析仪上应用本发明的方法分析挥发性物质

在和实施例 2 中相同的数据上，应用 MREM 方法，得到和实施例 2 中相同的结果。

MREM 方法应用参数：

锁峰区间：整个区间，m/z = 50~150（也就是没有指定是否用哪个特定的峰，不必指定特定的峰，让 MREM 方法自己找）。锁峰个数：每次只锁定一个。

优化方法：用于局部优化的模拟退火方法（上述文献③）。

优化次数：30 次。

当用 MREM 算法，在实施例 2 的数据中，用如上参数后，能够得到 30 个重建纯谱，将其中的相同的重建纯谱去除后，得到和实施例 2 一样的结果，也就是 4 个纯谱，得到的 TIC 结果也和图 18 一样。

实施例 4：MREM 和 tBTEM 算法连用，在气质联用分析仪上应用本发明的方法分析挥发性物质

在和实施例 2 中相同的数据上，连用 MREM 和 tBTEM 方法其中 MREM 算法的用处是：用其锁住区间的功能，和局部优化的功能；tBTEM 算法的用处是：用其可锁住多重区间的功能。得到和实施例 2 中相同的结果。

MREM 方法应用参数：

锁峰区间：两个区间，m/z = 50~100，和 m/z = 101~150（也就是利用 tBTEM 的锁住多重峰的功能，利用 MREM 方法的锁住区间的功能（不必指定特定的峰））。

锁峰个数：每个区间一个，共两个峰（这个就是 tBTEM 的特点）。

优化方法：用于局部优化的模拟退火方法（上述文献③）。

优化次数：30 次。

当连用 MREM 和 tBTEM 算法，在实施例 2 的数据中，用如上参数后，能够得到 30 个重建纯谱（重建纯谱包括纯的和复合谱），将其中相同的重建纯谱去除，并去除了复合谱后（详细算法参考上述文献②），得到和实施例 2 一样的结果，也就是 4 个纯谱，得到的 TIC 结果也和图 18 一样。

虽然，上文中已经用一般性说明及具体实施方案对本发明作了详尽的描述，但在本发明基础上，可以对之作一些修改或改进，这对本领域技术人员而言是显而易见的。因此，在不偏离本发明精神的基础上所做的这些修改或改进，均属于本发明要求保护的范围。

权 利 要 求 书

1、一种分析混合物成分的方法，包括:

（1）将混合物样品，利用色谱技术进行分离，得到初步的色谱图;

（2）对步骤（1）得到的初步的色谱图中的任意一个区间，根据需要进行采样，得到一系列在不同保留时间的谱图，称之为混合物谱图;

（3）利用熵最小算法的系列方法，从步骤（2）得到的混合物谱图中，计算得到各重建纯谱及其对应的伪浓度。

2、根据权利要求1所述的分析混合物成分的方法，其特征在于，步骤（1）中，得到的初步的色谱图中，包括纯成分的峰和混合物的重叠峰。

3、根据权利要求3所述的分析混合物成分的方法，其特征在于，对于混合物的重叠峰，利用步骤（2）和步骤（3）进行分析。

4、根据权利要求1~3任意一项所述的分析混合物成分的方法，其特征在于，步骤（1）得到的初步的色谱图中的任意一个区间，是指初步的色谱图的一段，或者全部。

5、根据权利要求1~4任意一项所述的分析混合物成分的方法，其特征在于，步骤（1）中，色谱技术为：气相色谱技术、液相色谱技术或者两者的组合。

6、根据权利要求1~5任意一项所述的分析混合物成分的方法，其特征在于，步骤（1）中，利用色谱技术进行分离，经检测器得到混合物的谱图。

7、根据权利要求6所述的分析混合物成分的方法，其特征在于，所述检测器选自：质谱、红外、紫外、荧光、分光光度仪和核磁共振检测器中的一种或几种的组合。

8、根据权利要求1~7任意一项所述的分析混合物成分的方法，其特征在于，步骤（3）中，熵最小算法的系列方法包括BTEM、tBTEM或MREM方法中的一种或几种的组合。

9、根据权利要求8所述的分析混合物成分的方法，其特征在于，利用BTEM或者tBTEM方法，每次运算，得到一个重建纯谱；利用MREM方法，每次运算，得到多个重建纯谱。

10、根据权利要求1~9任意一项所述的分析混合物成分的方法，其特征在于，步骤（3）中，对于含有未知物的混合物样品的分析，利用步骤（2）得到的混合物谱图，利用熵最小算法的系列方法，在不同的参数下，反复多次计算后，得到重建纯谱。

11、根据权利要求1~9任意一项所述的分析混合物成分的方法，其特征在于，步骤（3）中，对于含有标的物/已知物的混合物样品的分析，基于标的物/已知物的标准谱图，利用熵最小算法的系列方法计算，构建使用不同参数的重建方法，得到该标的物/已知物的重建纯谱；在多次的重建中，如果重建纯谱和标的物/已知物的标准谱图相比较是一致的，而且该重建纯谱的伪浓度是有意义的，则该混合物中含有该标的物/已知物，反之则不含有。

12、根据权利要求1~11任意一项所述的分析混合物成分的方法，其特征在于，在步骤（3）后将得到的重建纯谱和标准数据库中的谱图进行比对，确认它们的信息；若比对后重建纯谱和标准数据库中的某谱图一致，则说明重建纯谱代表的是标准数据库中的某谱图所代表的成分。