

谈谈弥散函数和“月份”基组

文/Sobereva @北京科音

First release: 2012-Jan-13 Last update: 2023-Sep-9

如果你对基组了解甚少，强烈建议阅读本文前先阅读《谈谈量子化学中基组的选择》(<http://sobereva.com/336>)。

1 何时需要弥散函数？

量子化学计算中的弥散函数是指指数很小的基函数，有很广的空间分布范围。加弥散函数的必要性，笔者根据大量理论计算文章和实践经验总结如下：

- 不加弥散结果一定没法用，必须加弥散：

计算偶极矩、多极矩

计算极化率、超极化率

计算里德堡激发态

计算阴离子体系能量、电子亲和能

- 加弥散很重要，强烈建议加弥散：

计算弱相互作用能（如果用 4-zeta 档次基组，而且不牵扯阴离子，那么不带弥散也完全可以）

计算红外、Raman、ROA 强度

- 加弥散对改进结果有益，计算条件允许可以加弥散：

计算反应势垒（仅对于 3-zeta 及以下档次基组而言。对于 def2-TZVP/def2-TZVPP，加弥散对势垒结果不会有明显改进）

优化阴离子体系或弱相互作用体系的结构、对其进行振动分析

如果体系只有部分原子带十分显著的负电荷或者其电子十分容易被极化，或者弱相互作用只涉及到体系中的一部分，限于计算能力没法给所有原子加弥散的话，那么只给这些区域的原子加上弥散函数就够了。

当弥散函数必不可少时，为了能加上弥散函数宁可降低 zeta 数，比如计算偶极矩、极化率时宁可从 3-zeta 降到 2-zeta 也要把弥散函数加上。但对于弥散函数不起到关键性作用但是加它有益处的时候，那么至少先把基组用到不低于 3-zeta 的档次再考虑加弥散，这样才最划算，因为此时增加基组价层分裂数带来的益处明显比加弥散函数大得多。

对于中性体系且其中不涉及显著的弱相互作用、没有局部带显著负电荷的时候，计算其上述未提到的特征，比如原子化能、价层电子激发、键能、电离能、几何优化、NMR 等等，完全没必要加弥散，否则对结果精度没什么改进，还徒增巨额计算时间、导致很多问题。记住乱用弥散函数只会自取其辱！

PS：笔者时不时看到居然有人算阳离子的普通问题（如能量、几何优化）时还加弥散函数，简直难以理喻，阳离子又不像阴离子那样有被束缚得较弱的电子，加什么弥散函数！？审稿人一看就知道这作者缺乏计算化学常识，好感度骤降！（后来我发现某些小白给阳离子加弥散的一个可能原因：有一次有人在我的群里贴了张讲基组的中文幻灯片里的图，背景是深蓝的，里面说对离子体系应该加弥散，居然也不区分是阳离子还是阴离子，这 ppt 真是超级误人子弟！似乎这 ppt 在中文的低水平的计算化学讨论场所流传还挺广，导致坑害了一堆初学者。所以千万别随便信网上来路不明的中文资料，这在《谈谈量子化学如何正确地入门》<http://sobereva.com/355> 里专门说了）

由于给氢乱用弥散函数的情况特别普遍，很多人还屡教不改，所以这里专门强调一下。给氢加弥散函数绝大多数情况没有任何意义。因为氢本身只有一个电子，电负性又小，在分子环境中一般都会被吸走不少电子，给它加弥散函数显然对结果不会有可察觉的改进，而有机体系里氢的数目通常很多，因此给氢加弥散还令计算量增加不少。此外，凡是给氢加弥散函数的时候肯定也会给重原子加，重原子的弥散函数会很大程度延展到氢那里去，也一定程度上表现了氢的弥散函数起到的作用，因此更没必要给氢加弥散函数。算普通问题给氢加弥散函数是多余的这点早就被大量文章的测试所证实，也被内行研究者所广泛认同。所以，需要对体系加弥散函数时，强烈不建议用诸如 6-311++G**，而建议用 6-311+G**。对于 Dunning 系列基组要加弥散的情况，也建议用后文说的月份基组而不是 aug-cc-pVnZ。即便研究氢键键能也不需要给氢加弥散函数，测试表明这对结果的影响微乎其微。但对于计算(超)极化率、氢带明显负电（如 LiH）等情况，给氢加弥散函数确实有重要意义，此时需要给氢加弥散。

2 弥散函数造成的问题

加了弥散函数会使基函数的完备性增加，从原理上讲似乎总应当对结果有益无害，但在实际计算中却会导致如下问题：

计算量猛增。这是众所周知的，尤其是 cc-pVnZ 系列，往往加了 aug-后就算不动了，第四节将讨论怎么解决这个问题。

加入了弥散函数后 SCF 经常比未加时难收敛得多。

弥散函数的化学意义很差，与原子轨道缺乏对应关系，会对希尔伯特空间下的波函数分析方法造成严重不良影响。比如 Mulliken 电荷会变得极烂，Mayer 键级会颇不可靠。原因不难理解，比如 A 原子一大堆弥散函数延伸到 B 原子空间内，因此 B 附近的电子分布有一部分会被这些弥散函数所描述，那么 Mulliken 布居分析就会把很多本应属于 B 的电子划归到 A 原子上，导致 A 的电荷过负而 B 的过正。

虚轨道（即非占据轨道）的化学意义变得更含糊。尤其是弥散函数下的 Hartree-Fock 计算，虚轨道空间分布范围往往会变得特别广，导致前线轨道理论分析完全不再适用。

导致出现基函数线性依赖问题引起数值问题。不过，一般量化程序中都会自动检验基函数重叠矩阵的本征值来适当砍掉些基函数以解决这个问题。

可能因数值计算精度问题导致体系结构和波函数对称性下降。比如你优化个有对称性的体系，一开始的结构程序是能判断出实际点群的，但优化之后，程序有时只能判断出更低阶点群了。如果你当前用了弥散函数，那么 8 成是弥散函数导致的，去掉弥散函数往往就能维持住对称性。

如果原始基组完备性不高，却增加了过多弥散函数，则本应该由价层基函数表现的效应会转而被弥散函数所表现，在一些问题的研究中可能引起不合理的结果，这其实属于分子内 BSSE 范畴。一个例子是 JACS,128,9342 发现后 HF 结合某些弥散版本的 Pople 基组（如 6-31++G**）算出来的苯的稳定结构竟然是弯的，或者说本该稳定的平面结构却有虚频。这是因为这些 Pople 基组中没有后 HF 计算较依赖的更高角动量基函数（尤其是 f），延伸过去的较为弥散的 s 和 p 基函数为了能充分等效展现出碳的更高角动量基函数的效应而引起了结构的弯曲。

3 常见的含有弥散函数的基组

先说一下弥散函数的一般特征。每种角动量的弥散函数的指数小于基组中其它同等角动量函数的最小指数的数倍。各种基组中的弥散函数数目、所涉及的角动量不同。弥散函数一般是非收缩的。

由于弥散函数对许多情况很重要，所以主流的基组大多数都有带弥散函数的版本。带弥散函数的版本有些是原基组的作者搞出来的，也有的是其它研究者提出来的。下面列举一些常见的：

1 Pople 系列基组：只给重原子加上一层 sp（即指数相同的一层 s 和一层 p）弥散函数就在 G 前头添上一个加号，如 6-31+G*；若同时还给氢、氦原子加上一层 s 弥散函数就在 G 前头再添上一个加号，如 6-311++G(2df,2p)。Pople 系列基组有很多，但是弥散函数的指数都是共用的，并未单独优化。（值得一提的是，6-311G*的耗时是要低于 6-31+G*的，对于弥散函数不起到关键作用时，前者的结果还比后者好。无数人用 6-31+G*的时候殊不知，对于他们所研究的问题，用 6-311G*划算得多得多！）

2 Dunning 相关一致性基组（cc-pVnZ 系列）：加上弥散函数的版本是 aug-cc-pVnZ 系列（aug=augmented），是给相应的 cc-pVnZ 基组的每种角动量函数上都添加一层同等角动量的弥散函数。例如 cc-pVTZ 对于 C 是 4s,3p,2d,1f，因此 aug-版本会增加一层 s、一层 p、一层 d 和一层 f 弥散函数。而 cc-pVDZ 对于氢是 2s,1p，因此 aug-版本会给它增加一层 s 和一层 p 弥散函数。

与 aug-cc-pVnZ 完全相同的弥散函数加到考虑描述核相关的 cc-pCVnZ 和 cc-pwCVnZ 基组上成为 aug-cc-pCVnZ 和 aug-cc-pwCVnZ；加到适合 DKH 计算的 cc-pVnZ-DK 上成为 aug-cc-pVnZ-DK；加到 cc-pV(n+d)Z 系列（这套基组是在 cc-pVnZ 基础上加入一层比较紧的 d 函数以改善外推时的收敛性）成为 aug-cc-pV(n+d)Z，其中仅 d 弥散函数的指数经过了重新优化。相关一致性赝势基组 cc-pVnZ-PP 也有弥散函数的版本 aug-cc-pVnZ-PP，弥散函数加入的类型和数量同 aug-cc-pVnZ 系列，但指数都重新优化了。另外，还可以对 cc-pVnZ 系列每个角动量加上多层弥散函数，d-aug-cc-pVnZ、t-aug-cc-pVnZ 就是给每个角动量上分别增加两层和三层弥散函数，极为昂贵，一般用于精确计算里德堡激发态、超极化率等目的。

3 Ahlrichs 的 def2-系列基组：目前这类基组并没有官方的带弥散函数的版本，这不能不说是个遗憾，但是有其它许多方式加弥散，见《给 ahlichrs 的 def2 系列基组加弥散的方法》(<http://sobereva.com/340>)。

4 Jensen 的极化一致性基组 pc-n：在 JCP,117,9234 中 Jensen 提出了给他的 pc-n 系列基组增加弥散函数的方法，弥散函数加到多高角动量可以根据需要自行选择。结果表明加了 s 和 p 弥散能让 DFT 计算的电子亲合势精度大有

提高，而进一步改善响应属性的计算结果还同时需要加更高角动量的弥散函数。

5 Lanl 赝势基组：在 JPCA,105,8111 中，作者给主族 Lanl2DZ 添加了一层 d 极化和一层 p 弥散函数，命名为 LANL2DZdp，在计算电子亲和能、振动频率、键长方面都比 Lanl2DZ 大有提高。LANL2TZ+和 LANL08+分别是给 LANL2TZ 和 LANL08 对于第一周期过渡金属增加了一层弥散，这是考虑到 d 壳层充满的第一周期过渡金属有时容易被极化。

除了 LANL2TZ+和 LANL08+等基组的弥散函数的指数用的是 even-tempered 方式推导出的以外，上面提到的弥散函数的指数大多都是来自于不同方式最小化阴离子计算的能量（注：原基组的收缩系数和指数保持不变，弥散函数只是单纯地累加到原基组）。然而，计算能量好的弥散函数基组用于计算其它属性，尤其是强烈依赖弥散函数的（超）极化率未必好，或者说性价比不高。有不少人为了能让（超）极化率等响应属性在较低的计算成本下有满意的计算结果，令弥散函数，乃至整个基组都直接来自于优化响应属性的计算，这里也举几个例子：

Sadlej POL：也叫 Sadlej pVTZ 基组，从 1988 年陆续提出，参数是优化计算极化率得到的。大小接近 cc-pVTZ，算极化率精度则接近昂贵得多的 aug-cc-pVTZ。

Sadlej ZPOL：2004 年陆续提出。对 POL 基组进行简化。适合计算很大体系偶极矩和极化率，和 6-311+G*耗时相仿佛，但算极化率精度比 6-311++G(2df,2p)都好。

Sadlej LPol-ds：2009 年提出。LPol 系列基组分为 LPol-ds/dl/fs/fl，依次增大。LPol-ds 是最小的，但也比 POL 大得多得多，算第一超极化率精度极好，和 d-aug-cc-pVTZ 相仿佛，耗时则是同档次精度中最低的。只对 C、H、O、N、F 有定义。

def2-SVPD、TZVPD、TZVPPD、QZVPD、QZVPPD 基组：在 JCP,133,134105 中提出，是分别对 def2-系列的 SVP、TZVP、TZVPP、QZVP、QZVPP 基组加上弥散函数。弥散函数的指数通过优化原子的 HF 极化率得到。

LFK 赝势基组：是在 SBKJC 赝势基组基础上对 39 个主族元素增加弥散和极化函数，目的是在赝势基组的计算量下达到全电子 Sadlej 基组下的极化率计算精度。指数和收缩系数是通过逼近实验原子极化率或高精度相对论计算值得到的。

如果某种基组目前没有人提出带弥散函数的版本，而计算却又需要弥散函数，那么可以考虑将其它尺寸差不多的基组的弥散函数挪过来用。也可以通过 even-tempered 方式基于现有基组来构建，另外还有人提出可以将基组中指数最小的 s 和 p 的指数除以 3 作为 s 和 p 弥散函数的指数，见 JCTC,7,3027。虽说这样算出来的指数对特定问题肯定没专门优化出来的指数那么好，但对于在优化指数过程中没考虑到的更广泛的体系和问题的类型，直接算出来的指数也有可能表现得更好。更多讨论见前面提到的《给 ahlich 的 def2 系列基组加弥散的方法》。

4 降低弥散函数的使用量以节约时间

Pople 系列基组和 cc-pVnZ 系列基组是目前最流行的基组。从前面的介绍可知，后者的带弥散函数的版本额外加入的函数量比 Pople 系列基组的带弥散的版本多得多，因为包含了更高角动量壳层（同时注意，角动量越大的壳层内包含的基函数数目越多，且积分越耗时），此外后者一律给氢加上弥散函数，而不是像 Pople 系列基组那样通过 +和++来区分。根据 JCTC,7,10 的统计，一般应用中 aug-cc-pVnZ 基函数数目比 n 相同的 cc-pVnZ 基组多了一半有余，在 Gaussian 下的 MP2 计算中，前者计算耗时是后者的 2.4 倍（n=D），5 倍（n=T）和 6.3 倍（n=Q）。aug-往往带来过大的计算负担而使研究者难以承担，还使收敛更困难。一些研究者发现高角动量弥散函数的价值并不大，因此考虑能否适当删掉一些高角动量的弥散函数来缓解这些问题。不同的人提出了不同的删减 aug-cc-pVnZ 弥散函数的方法。但是不同的删减方法往往只是在个别应用中顺带进行讨论，缺乏系统性。Truhlar 等人在 JCTC,7,10 和 JCTC,7,3027 中给出了一套系统的删减方法，由不同的删减程度所得到的“月份”系列基组弥补了 cc-pVnZ 和 aug-cc-pVnZ 之间的空白，下面将具体介绍。

首先说一下 minimal augmentation(maug)概念，它是指对一个没有弥散函数的基组加入最低限度的弥散函数，具体来说，就是只给重原子加一层 s 和 p 弥散函数，而不给氢加任何弥散函数。maug-cc-pVnZ 基组，就是将 aug-cc-pVnZ 的重原子的 s 和 p 弥散函数挪到 cc-pVnZ 上，或者说，是把 aug-cc-pVnZ 的氢的所有弥散函数和重原子的角动量高于 p 的弥散函数都砍掉的结果。

月份基组像 maug-基组一样砍掉所有氢的弥散函数，但是重原子的高角动量弥散函数不是都砍掉，而是根据月份顺序依次砍掉最高的，直到等价于 maug-基组为止：

jul-cc-pVnZ：将 aug-cc-pVnZ 的氢的所有弥散砍掉。重元素的弥散函数不变。

jun-cc-pVnZ：将 jul-cc-pVnZ 的重元素的角动量最高的弥散函数砍掉。

may-cc-pVnZ: 将 jun-cc-pVnZ 的重元素的角动量最高的弥散函数砍掉。

apr-cc-pVnZ: 将 may-cc-pVnZ 的重元素的角动量最高的弥散函数砍掉。

...

比如对于 aug-cc-pVQZ 下的碳原子, 最高角动量是 g, 所以 jul-、jun-、may-和 apr-分别对应于: 不砍掉任何弥散函数; 砍掉 g 弥散函数; 砍掉 f,g 弥散函数; 砍掉 d,f,g 弥散函数。由于 apr-cc-pVQZ 已经等于 maug-cc-pVQZ, 所以 n=Q 时的月份基组就没有 mar-cc-pVQZ 了。如果是 aug-cc-pV5Z, 由于最高角动量增加了 1, 所以相应的月份基组的下限也会降一个月成为 mar。

之所以月份基组是从 jul 开始, 是因为完整的基组开头的 aug-可看作 august。月份基组只是从中简单地删除函数, 其它的函数的指数和收缩系数并没有重新做优化。

众所周知, HF/KS-DFT 这样的基于单电子有效势的方法对于基组的角动量要求并不高, 需要有弥散函数的情况下, 通常使用 maug-级别就够了。aug-级别里的高角动量弥散函数给 HF/KS-DFT 用实在太浪费。maug-cc-pVDZ 是推荐用于大体系的基组, 而精度要求高一些的话, 用 maug-cc-pVTZ (等价于 may-cc-pVTZ) 就已经很好了。

后 HF 计算对于高角动量函数比 HF/KS-DFT 敏感得多, maug-的弥散函数级别显得偏小了。对于必须需要弥散函数的情况, 中低精度的计算建议用 jul-cc-pVDZ, 中等精度的计算建议用 may-cc-pVTZ, 中高精度计算建议用 jul-cc-pVTZ, 高精度的计算建议用 jun-cc-pVQZ。之所以 zeta 数越多, 所需精度越高, 却可以将月份适当降低, 是因为随着价层分裂数目的增加, cc-pVnZ 中最外层的基函数越弥散, 越能等效起到弥散函数的作用, 相同 zeta 数下不同级别的月份基组的精度差异越小。值得一说的是, 非常需要弥散函数的任务中, 给 cc-pVnZ 哪怕仅加入 maug-级别的弥散对结果的改进都要比提高 zeta 数要显著得多, 即弥散函数该有的时候必须有, 但不需要角动量太高。

可以说, 只要适当使用月份基组, 就没必要再用 aug-cc-pVnZ 基组了。可能有人觉得, 将氢原子的全部弥散函数去掉是不是有点过分了, 实际上这没有问题。因为氢原子电负性小, 电子并不富集、弥散, 它周边的重原子的弥散函数也能起到补充作用。就连高精度弱相互作用测试集 S66 在外推完备基组下的结果时都删去了氢的弥散函数, 更何况一般的计算。但对于个别情况, 比如 LiH, 由于 Li 的电负性很小, 氢会带有很大负电荷, 这时候氢也得加弥散。

正如这一节强调的, 氢的弥散函数绝大多数情况没什么用, 所以很多人用比如 6-311++G**的时候都应该改为 6-311+G**, 这样能节省大量时间又不会使精度有什么损失。

5 月份基组在 Gaussian 中的使用方法

从 Gaussian09 D.01 版开始, 直接支持月份基组。直接写比如 may-cc-pVTZ、jul-cc-pVQZ 这样就可以了。另外, Gaussian 里还有个 spAug-cc-pV*Z 的写法, 也就是相对于 cc-pV*Z 给重原子加上 s 和 p 弥散(相当于 maug-cc-pV*Z), 但是还给氢加上 s 弥散。

如果用的是 Gaussian 老版本, 就得通过自定义基组方式来使用。这里以对甲烷使用 may-cc-pVTZ (等价于 maug-cc-pVTZ) 为例进行说明。

进入 BSE 基组库 (<https://www.basissetexchange.org>), 周期表中点击碳, 在左侧列表里选 aug-cc-pVTZ, 然后 Format 选 Gaussian94, 然后点 Get Basis Set 按钮, 弹出的窗口有如下信息

```
****
C      0
S      8      1.00
      8236.0000000      0.0005310
      1235.0000000      0.0041080
      280.8000000      0.0210870
      79.2700000      0.0818530
      25.5900000      0.2348170
      8.9970000      0.4344010
      3.3190000      0.3461290
      0.3643000      -0.0089830
S      8      1.00
      8236.0000000      -0.0001130
      1235.0000000      -0.0008780
      280.8000000      -0.0045400
      79.2700000      -0.0181330
      25.5900000      -0.0557600
      8.9970000      -0.1268950
      3.3190000      -0.1703520
      0.3643000      0.5986840
S      1      1.00
      0.9059000      1.0000000
```

```

S 1 1.00
0.1285000 1.0000000
S 1 1.00
0.0440200 1.0000000
P 3 1.00
18.7100000 0.0140310
4.1330000 0.0868660
1.2000000 0.2902160
P 1 1.00
0.3827000 1.0000000
P 1 1.00
0.1209000 1.0000000
P 1 1.00
0.0356900 1.0000000
D 1 1.00
1.0970000 1.0000000
D 1 1.00
0.3180000 1.0000000
D 1 1.00
0.1000000 1.0000000
F 1 1.00
0.7610000 1.0000000
F 1 1.00
0.2680000 1.0000000

```

这些是 aug-cc-pVTZ 中对碳原子的弥散函数的定义，没字母的那些行中第一列是指数，第二列是收缩系数。may-是砍掉两个最高角动量弥散函数，所以 d 和 f 弥散函数就不要了，只保留 s 和 p 的弥散函数。同等角动量中指数最小的基函数就是弥散函数，因此

```

D 1 1.00
0.1000000 1.0000000

```

和

```

F 1 1.00
0.2680000 1.0000000

```

应该被舍弃。

may-cc-pVTZ 下计算 CH4 分子的 Gaussian 输入文件应为

```

#P wb97xd/gen
Divokej Bill
0 1
C -2.32704417 -0.72320845 -0.00946732
H -1.97038974 -1.73201845 -0.00946732
H -1.97037133 -0.21881026 0.86418419
H -1.97037133 -0.21881026 -0.88311882
H -3.39704417 -0.72319527 -0.00946732
C 0
S 8 1.00
8236.0000000 0.0005310
1235.0000000 0.0041080
280.8000000 0.0210870
79.2700000 0.0818530
25.5900000 0.2348170
8.9970000 0.4344010
3.3190000 0.3461290
0.3643000 -0.0089830
S 8 1.00
8236.0000000 -0.0001130
1235.0000000 -0.0008780
280.8000000 -0.0045400
79.2700000 -0.0181330
25.5900000 -0.0557600
8.9970000 -0.1268950
3.3190000 -0.1703520
0.3643000 0.5986840
S 1 1.00
0.9059000 1.0000000
S 1 1.00
0.1285000 1.0000000
S 1 1.00
0.0440200 1.0000000
P 3 1.00
18.7100000 0.0140310
4.1330000 0.0868660

```

```

1.2000000      0.2902160
P  1  1.00
0.3827000      1.0000000
P  1  1.00
0.1209000      1.0000000
P  1  1.00
0.0356900      1.0000000
D  1  1.00
1.0970000      1.0000000
D  1  1.00
0.3180000      1.0000000
F  1  1.00
0.7610000      1.0000000
****
H  0
cc-pVTZ
****

```

这里对 H 使用 cc-pVTZ，是由于月份基组对于氢来说相当于砍掉了所有 aug-cc-pVnZ 基组的弥散函数，和 cc-pVnZ 是等价的。

另外，BSE 上目前也有月份基组的定义可以直接用。比如在 BSE 界面里点击碳，直接找 may-cc-pV(T+d)Z 来用即可，这和我们上面自行修改得到的相同。