科学通报
第40卷 第8期 CHINESE SCIENCE BULLETIN 1995年4月

模拟退火及紫外光谱法用于氨基酸多组分分析*

李志良 曾鸽鸣 王树信 梁本熹 木村敏① 吉田英①
(湖南大学化学化工系，长沙 410082; ①日本国立丰桥技科大学，丰桥〒 441)

关键词 模拟退火算法 多元分析 氨基酸 光谱分辨

模拟退火(SA)系寻找全局最优并能跨越局部最优的随机优化算法，它源于对高温物质的退火过程的模拟即在给定温度下对微观粒子(如原子)平衡的统计力学模拟. SA[1]采用模拟算法及随机抽样; Kirkpatrick 等[2]深入研究了组合优化问题; Bohackevsky 等[3]提出了通用模拟退火(GSA)法; Kalivas 等[4, 5]研究了 GSA 用于多元校正. 本文将 SA 与 GSA 用于多组分分析，获得良好效果.

1 原理与方法

1.1 退火过程及模拟

固体物质处于熔融状态，所组成的微观粒子处于完全随机排列组合，以足够低的降温速率退火，维持体系在各温度下微粒达平衡，则体系服从 Boltzmann 分布:

$$f_i=g_i\exp(-E_i/T_i)=[1/h(i)]\exp(-E_i/bT_i), \quad (1)$$

此处 i 指状态，T_i 为温度，E_i 为能量，g_i 为常数，$h(i)$ 为分函数，b 为 Boltzmann 常数; f_i 为出现概率. 随温度降低，高能状态出现概率越来越小，最终趋于零. 微粒状态按 Boltzmann 分布趋于能量最低的状态即基态. 用 Monte Carlo 模拟给定温度下微粒达热平衡的过程，对微粒当前态随机地进行微扰. 令 E_a 和 E_b 分别为当前和微扰新状态的能量，能差为 $\Delta E=E_b-E_a$. 若 $E_a>E_b$, 则无条件接受新状态并代替当前态; 若 $E_a\leqslant E_b$, 则此概率 $\exp[-(E_b-E_a)/T]$ 接受此不利状态为当前态. 继续徐徐降温并重复抽样，直至获最低能量状态.

1.2 模拟算法

将状态对应於待优化参数 x; 对能量 E 相当于目标函数 J; 视退火温度 T 与 Boltzmann 常数为控制随机搜索程度的参数 B. 具体步骤如下: (1) 设定初始态 x_a 并计算其目标函数 $J_a=J(x_a)$. (2)施加随机微扰得新状态 x_b, 计算 $J_b=J(x_b)$ 及能差 $\Delta J=J_b-J_a$. (3) 比较 J_a 与 J_b, 若 $J_a>J_b$, 则无条件接受 x_b 为当前态，若 $J_a<J_b$, 则算概率 $f=\exp(-B\Delta J)$. (4) 产生一均匀随机数 $u\in[0, 1]$; 若 $f\geqslant u$, 则接受 x_b 并作为当前态，否则拒绝. (5)重复微扰直至抽样达稳定即收敛得全局最优解. (6)若拒绝则重新依上计算. 适择控制参数 B[2, 6]便可达全局最优.

1994-04-20 收稿，1994-12-14 收修改稿
*日本政府文部省与国家教委及国家自然科学基金资助课题

1.3 通用算法

将 $J=J(x)$ 定义为多维连续函数，并寻优得某状态 x_0 使 $J(x_0)$ 达最优．设当前态 x_a，微扰产生新随机态

$$x_b=x_a+v\cdot s, \tag{2}$$

v 为方向矢量，s 为变化步长．任意产生 n 个 $N(0,1)$ 随机数 $W_i(i=1,2,\cdots,n)$，计算其余弦方向 $v_i=w_i/\sqrt{\Sigma w_i^2}$，则状态为

$$x_b=x_a+sv\Rightarrow x_{bi}=x_{ai}+v_i\cdot s. \tag{3}$$

GSA 取消了 SA 改变 B 的控温循环，并修正接受概率：

$$f=\exp[-B\cdot\Delta J\cdot(DJ)^a]=\exp[-B\cdot\Delta J/DJ], \tag{4}$$

式中 B 仍为控制因子，$\Delta J=E_b-E_a$ 为微扰能差，$DJ=E_a-E_b$ 为相对能差，$a<0$ 为任意负值，当 $a=0$ 时 GSA 还原为 SA．常取 $a=-1$，则接受概率降低，有利收敛．为搜索接近全局最优时应不再接受不利状态(即接受概率近零)．步长 s 在 GSA 中通常不变，也可变化[7]．本文逐步降低步长 s，以提高优化精度．J_0 为全局最优点的目标函数，若已知可直接选取；若未知或难预先确定，则可先尝试某值再据运算予调整．故 GSA 提高了优化精度和收敛速度．

1.4 控制参数 B

曾有建议[3]适选 B 值使接受概率 $f\in[0.5, 0.90]$，常使 $f=0.8$．若选择不当则或使算法 f 过低而陷入局优；或使算法 f 过高而难於收敛，即完全随机游走．

1.5 多元校正

最常见的多组分光谱分析定量模型参见文献[8]．由纯组分直接测定或标准样处理间接确定量测吸光系数．

1.6 多元定量

测量未知试样的吸收光谱数据可依 SA 及 GSA 法预测未知浓度，即实现多组分同时定量[8]．

2 实验部分

2.1 仪器与试剂

日本岛津 UV-240 和美国 Beckman DU-7HS 及国产上海 752 型紫外可见分光光度计；IBM-PC/AT 386 微机及兼容机；VAX/DMS 工作站用於编制运行 SA 和 GSA 及扫描算法程序．色氨酸 Trp，酪氨酸 Tyr，苯丙氨酸 Phe，及 3,4- 二羟基苯丙氨酸 Dhp 等均为色谱纯试剂(Merck)，用常规法配成 0.1 mol/L HCl 储备液并稀释为工作液，其它试剂均为 AR 以上．

2.2 实验方法

配成纯组分标液和多组分混液测量紫外光谱，扫描波长范围为 190～300 nm，狭缝宽度为 1.0 nm，在 200～290 nm 范围内间隔 1 nm 读取数据输入计算机处理．

3 结果与讨论

3.1 紫外光谱

四组分氨基酸光谱(图 1)重叠严重，须予分离定量或作多元分辨，方能实现其多组分分析．

3.2 算法参数影响

以对称函数 $f(x_1, x_2)=x_1^2+2(1-\cos2\pi x_1)+3x_2^2+4(1-\cos4\pi x_2)$ 作为目标函数为例，考察 GSA

中循环步长 s 及控制参数 B 对求解精度的影响. 该函数有若干局部最优点和零值全局最优点 $(x_1=0,\ x_2=0)$. 假定优化约束区域为 $-1\leqslant x(x_1,\ x_2)\leqslant 1$, 任取点 $(x_1=1,\ x_2=1)$ (实为最远者的四点之一) 为初始点作最优搜索. 取 $B=1.0$, 步长固定为 $s=0.2$, 经 3 次寻优得目标函数即搜索误差分别为 7.3%, 4.2%, 10.5%; 改变搜索步长 (以半幅收缩), 仍取 $B=1.0$, 步长初值为 $s=0.2$, 以半幅收缩, 经 3 次寻优搜索得目标函数为 0.072 0‰, 0.038 0‰, 0.002 9‰. 表明引入收缩步长后, 其优化求解精度大大改善. 参数 B 对求解结果亦有较大影响, 若 B 过小 (如 $B=0.01$), f 将过大, 抽样不能收敛; 若 B 太大 (如 $B=100$), 则 f 将太小, 算法易陷入局部最优而得不到全局最优. 通常算法受到许多因素的制约, 比如局部最优点数、函数形状、变化缓急等. 欲使算法在极值点附近搜索实际上也难於实现. 为保证能搜索整个区域, 步长 s 宜取稍大. 对有约束优化, 步长宜取约为区域 10% 左右. 倘欲改善精度, 则步长可逐渐收缩.

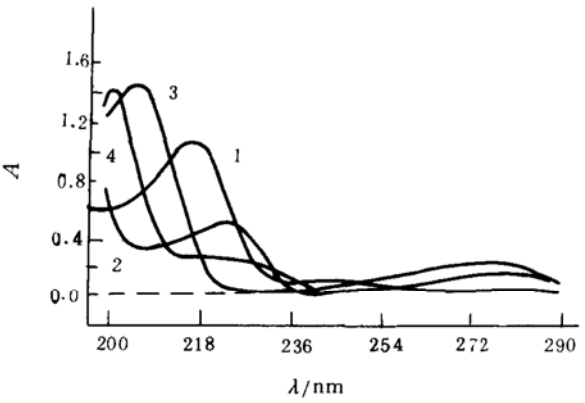

图 1　紫外光谱

氨基酸: 1 为 Trp, 2 为 Tyr, 3 为 Phe, 4 为 Dhp. 介质为 0.1 mol/L HCl

表 1　四组分混合氨基酸试样同时测定结果

试样号	加入量 Added($\times 10^{-6}$)				测得量 Found($\times 10^{-6}$ 上排数据) 及回收率 Recovery/%(下排数据)											
					卡尔滤波				因子分析				本法 GSA			
	Tyr	Trp	Phe	Dhp	Tyr	Trp	Phe	Dhp	Tyr	Trp	Phe	Dhp	Tyr	Trp	Phe	Dhp
1	4.670	5.180	14.33	8.280	4.570	5.127	14.01	8.274	4.501	5.214	14.572	8.532	4.568	5.126	14.012	8.276
					97.85	98.97	97.76	100.8	96.4	100.7	101.7	103.9	97.82	98.96	97.78	100.82
2	3.736	6.216	23.89	4.104	3.734	6.137	23.43	4.116	3.716	6.225	23.792	4.306	3.726	6.231	23.401	4.126
					99.95	98.73	98.05	100.3	99.5	100.1	99.6	104.9	99.73	100.24	97.95	100.54
3	3.736	4.144	33.45	12.31	3.561	4.159	32.71	12.43	3.422	4.225	32.600	12.872	3.561	4.168	32.70	12.54
					95.32	100.4	97.80	100.9	91.6	101.9	97.5	104.6	95.32	100.58	97.76	100.87
4	8.406	2.072	23.89	6.156	8.315	2.072	24.74	6.428	8.161	2.161	23.953	6.626	8.308	2.072	24.47	6.428
					98.92	100.0	103.5	104.4	97.1	104.3	100.3	107.6	98.92	100.0	102.43	104.4
5	7.472	1.036	9.556	16.42	7.319	1.045	9.998	16.428	7.086	1.131	9.732	16.965	7.149	1.045	9.978	16.428
					97.96	100.9	104.6	100.9	94.6	109.2	101.8	103.3	95.68	100.9	104.42	100.9
6	6.538	2.072	4.778	10.26	6.530	2.026	4.516	10.34	6.500	2.125	5.014	10.341	6.530	2.026	4.512	10.342
					99.88	97.78	94.51	100.8	99.4	102.6	104.9	100.8	99.88	97.78	94.32	100.8
7	5.604	5.180	14.33	2.052	5.646	5.060	13.72	2.146	5.689	5.167	14.646	2.164	5.649	5.060	13.709	2.146
					100.8	97.68	95.75	104.6	101.5	19.7	102.2	105.4	100.80	97.7	95.67	104.6
8	2.902	4.144	23.89	12.31	2.774	4.075	23.62	12.40	2.783	4.202	23.954	12.612	2.774	4.070	23.62	12.526
					99.00	98.35	98.86	100.7	99.3	101.4	100.3	102.5	99.0	98.9	98.9	101.755
9	2.802	4.144	23.89	12.31	2.871	4.058	23.58	12.41	2.893	4.187	24.105	12.436	2.872	4.054	23.456	12.412
					102.5	97.93	99.14	100.8	103.3	101.0	100.9	101.0	102.5	97.83	98.18	100.8
10	4.670	3.108	23.89	2.052	4.890	3.021	24.43	1.995	4.941	3.005	24.702	1.986	4.890	3.021	24.507	1.989
					104.7	97.22	102.3	97.23	105.8	96.7	103.4	96.8	104.7	97.22	102.58	96.93

3.3 四组分氨基酸同时测定

对四组分氨基酸混合体系的光谱数据处理，参数 B 的选择无规可循且取值无界；步长 s 选取亦无经验可言但区域有下限，即约束条件 $x_i \geqslant 0$；我们依据测试光谱数据确定各组分浓度可能上限，然后以此区域边界取初始步长(约为 1/10)求解．均经百次试算初始不接受不利状态的概率 P，若 $P \geqslant 0.9$ 或 $P \leqslant 0.5$，则返回重设参数．以最小二乘准则依 GSA 预测未知试样中各氨基酸组分的浓度，与卡尔曼滤波(KF)[8, 9]及目标因子分析(FA)求解比较，结果相互一致，优化精度与预测结果良好：回收率为 95.3% ～ 104.6%.

3.4 初步结论

GSA 与 SA 系有用的化学计量学新算法，用於同时光谱测定，为氨基酸分析提供了一种优良的和有效的新方法．也可望在其它方面取得广泛应用．

参 考 文 献

1 Metropolis N, Rosenbluth A, Rosenbluth M *et al*. Equation of state calculations by fast computing machines. J Chem Phys, 1953, 21: 1 087～1 092

2 Kirkapatrick S, Gelatt Jr C D, Vecchi M P. Optimization by simulated annealing. Science, 1983, 220: 671～680

3 Bohackevsky I O, Jahnson M E, Stein M L. Generalized simulated annealing for function optimization. Technom, 1986, 28(3): 209～217

4 Kalivas J H, Roberts N, Sutter J M. Global optimization by simulated annealing with wavelength section for ultraviolet-visible spectrophotometry. Anal Chem, 1989, 61: 2 024～2 203

5 Kalivas J H. Generalized simulated annealing for calibration sample selection from an existing set and orthogonalization of undesigned experiments. J Chemom, 1991, 5(1): 37～48

6 Collins N E, Eglese R W, Golden B L. Simulated annealing-an annotated bibliography. Am J Math Mang Sci, 1988, 8(3/4): 209～307

7 Vanderbilt D, Louie S G. A Monte Carlo simulated annealing approach to optimization over continuous variables. J Comput Phys, 1984, 56: 259～271

8 Li Z L, Shi L M, Li M L *et al*. A highly sensitive and selective method for trace multicomponent analysis. Acta Chim Sin, 1990, 48(11): 1 101～1 107

9 Shi L M, Xu Z H, Li Z L *et al*. Application of the Kalman filter algorithm to the simultaneous determination four amino acids by direct uv spectrophyotometry. J Micronutr Anal, 1990, 8: 1～12