Decomposition-Based Algorithms Using Pareto Adaptive Scalarizing Methods

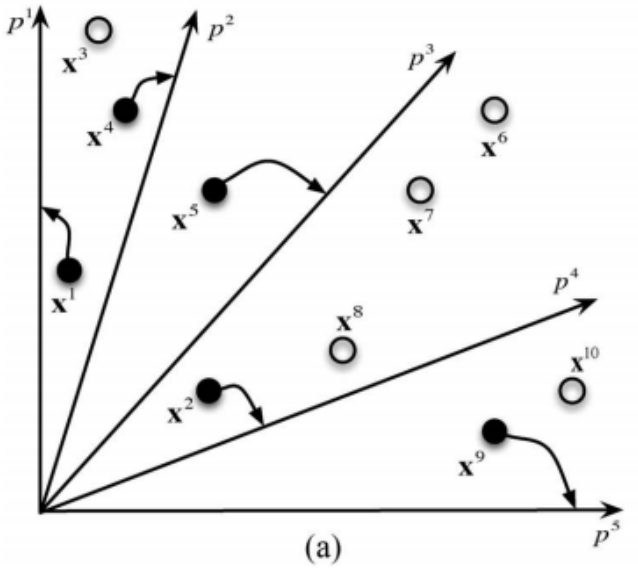
By 张谋钊

2017.4.27

MOEA/D

影响MOEA/D表现的两个部分

- Weights
  - ➢ evenly distributed
  - ➢ Adjusted adaptively during the search
- Aggregation function（待分析）



content

- 分析聚合函数
- P值的作用
- 提出PaS
- MOEA/D-PaS
- 实验
- 结果分析

分析Aggregation function

Weighted distance(also called Lp) scalarizing methods-- MOEA/D中较为常用的一类函数

$$g^{\mathrm{wd}}(\mathbf{x}|\mathbf{w},p)=\left(\sum_{i=1}^{m}\lambda_i\left(f_i(\mathbf{x})-z_i^*\right)^p\right)^{\frac{1}{p}},\ p\geq 1$$
$$\lambda_i=\left(\frac{1}{w_i}\right)^p \qquad (2)$$

两个特例：

当p=1时，为加权和（ weighted sum ）

$$\begin{aligned}&\text{minimize}\quad g^{ws}(\mathbf{x}|\mathbf{w})=\sum_{i=1}^{m}w_i f_i(\mathbf{x})\\&\text{subject to}\quad \mathbf{x}\in\Omega\end{aligned}$$

当p=∞时，为加权切比雪夫（ weighted Chebyshev）

$$\begin{aligned}&\text{minimize}\ g^{te}(x|\lambda,z^*)=\max_{1\leq i\leq m}\{\lambda_i|f_i(x)-z_i^*|\}\\&\text{subject to}\ x\in\Omega\end{aligned}$$

分析Aggregation function

分析该类方法，从两方面入手:

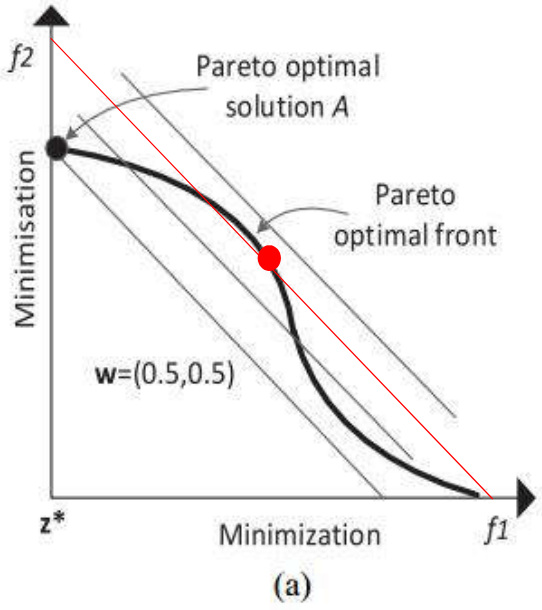
1)Search Ability

由聚合函数产生的向着真实PF的选择压力

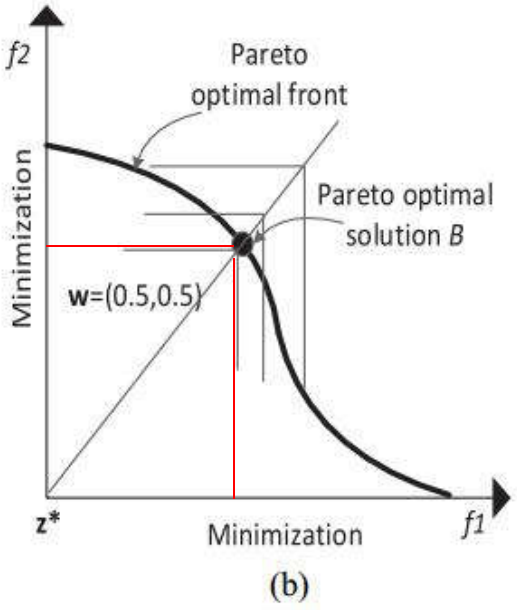
2) Robustness on PF Geometry

最后求得的PF与真实的PF在几何形状上的相似度

Search Ability



加权和
(WS)



加权切比雪夫
(TCH)

假设问题为最小化问题

定义等高线与坐标轴围成的区域为更优解区域

从图中可以看出WS的更优解区域是TCH的两倍

【结论】WS的搜索能力比TCH强

TCH的更优解区域其实为(1/2)^m，m为目标个数，所以随目标数增大，切比雪夫的搜索能力降低

Search Ability

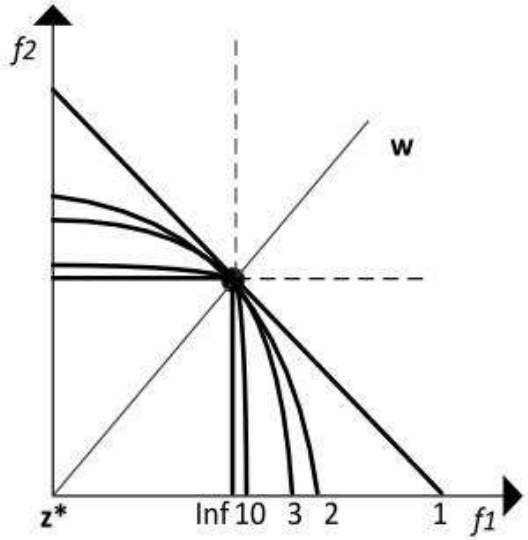


Fig. 3. Contour curves of the L_p scalarizing method with different p values.

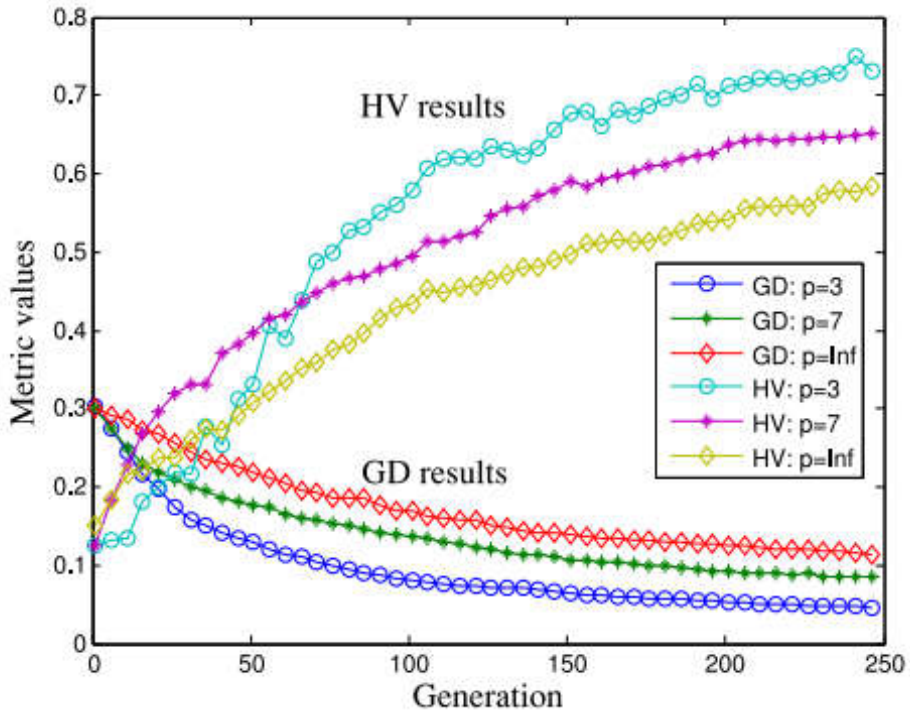
【结论】

1. p越大，等高线与坐标轴围成的面积越小，找到更优解的概率越小，所以搜索能力越差
2. 更优解面积的差异在接近真实PF时会缩小

【小实验】

用不同p值的聚合函数跑同样的测试集，除p以为参数均相同

【实验结果】



Search Ability

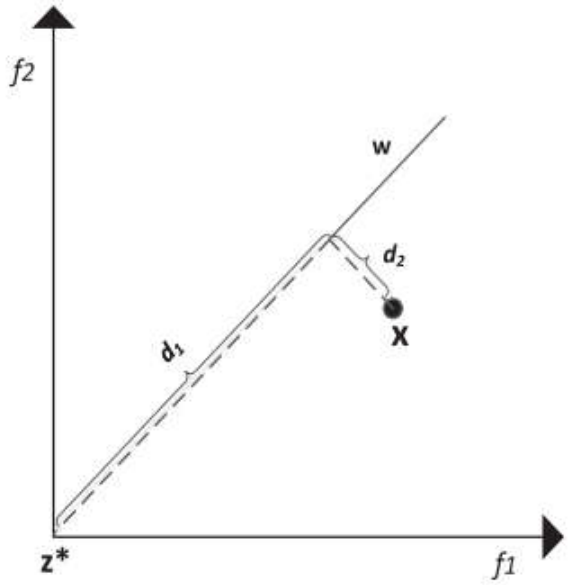


Fig. 5. Measurement of two distances in an L_p scalarizing method.

为更好分析p值对聚合函数的影响，
定义两个距离d1和d2，如图

- d1越小，解的收敛度越高
- d2越小，解的分布性越好

Search Ability

P值对d1和d2的影响：

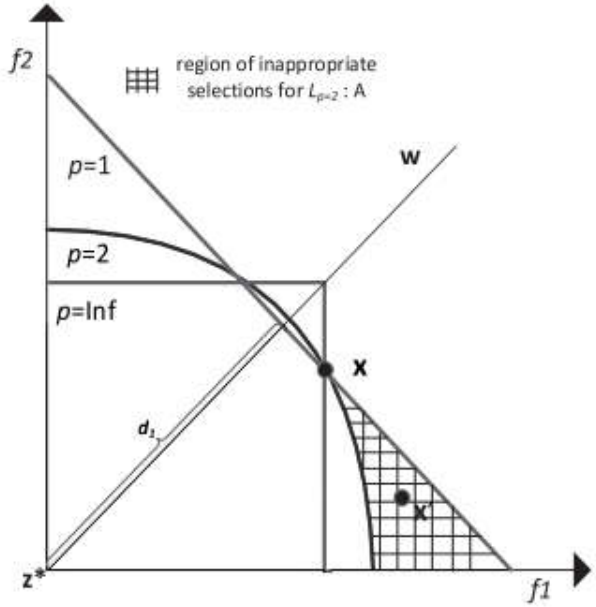


Fig. 6. Illustration of the influence of p to the d_1 distance.

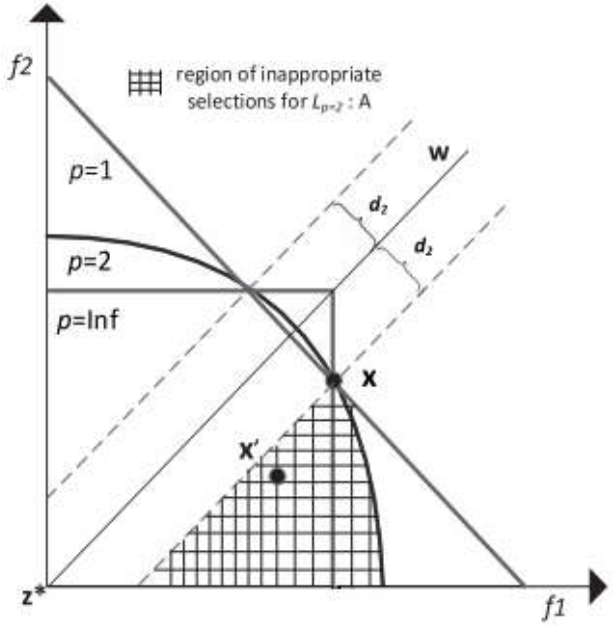


Fig. 7. Illustration of the influence of p to the d_2 distance.

【结论】p越大，所选解的d1越大，收敛性越好 ，反之相反。

【结论】p越大，所选解的d2越小，分布性越好，反之相反。

Search Ability

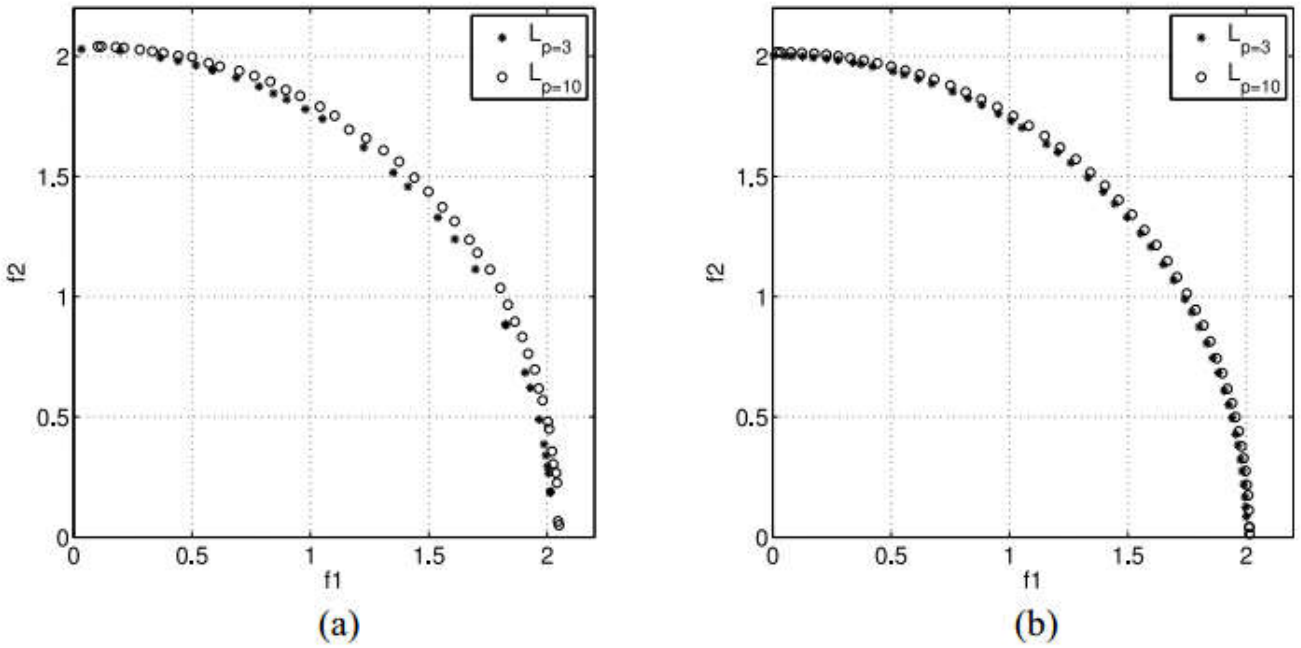
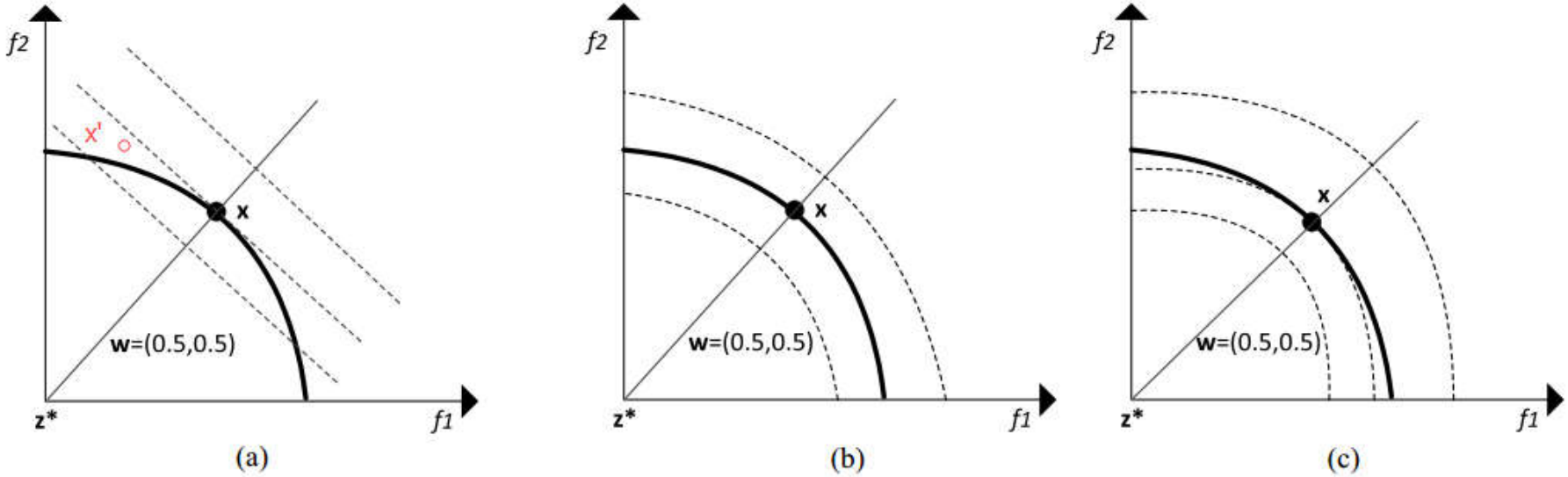


Fig. 8. Illustration of solutions obtained by MOEA/D using $L_{p=3}$ and $L_{p=10}$. (a) 100 generations. (b) 500 generations.

【结论】随着迭代的增加，p值得影响力越来越小。

Robustness of Different Lp Scalarizing Methods on the PF Geometry



【结论】

1. P值较小时，可能找不到PF为凸面的解。
2. 当p大于某个值时，或者说等高线的曲率大于PF的曲率时，该问题解决。

Robustness of Different Lp Scalarizing Methods on the PF Geometry

假设PF为:

$$a_1 f_1^{k_1}(\mathbf{x}) + \cdots + a_i f_i^{k_i}(\mathbf{x}) + \cdots + a_m f_m^{k_m}(\mathbf{x}) = 1$$

满足以下条件时，聚合函数有效（能够找到最优解）

$$g^{\mathrm{wd}}(\mathbf{x}^*|\mathbf{w}, p) < g^{\mathrm{wd}}(\mathbf{x}'|\mathbf{w}, p), \text{ for all } \mathbf{x}' \qquad (5)$$

根据PF和公式(5),解得：p > max(ki)

【结论】

1. 当p大于max(ki) ，聚合函数能够找到凸面PF的最优解
2. p越大，聚合函数找到的PF在几何方面越健壮。

P值的作用

总结：

- P越大，搜索能力越弱，分布性越好，最终得到的PF几何健壮性越强
- P越小，搜索能力越强，分布性越差，最终得到的PF几何健壮性越差
- 随着迭代的增加，两者的差异会慢慢减弱

PaS(Pareto Adative Lp Scalarizing)的提出

目标1：找到一个有效的且搜索力强的聚合函数，即p值

分析：

假设当前有解集合S={x1, x2, •••, xn}, 有n个权重向量，对应n个单目标子问题。在这种情况下，可通过以下公式得到p值

$$\begin{aligned} &\min \quad p, \quad p \in P \\ &\text{s.t.} \quad \forall \mathbf{x}^k : g^{\mathrm{wd}}\left(\mathbf{x}^* \middle| \mathbf{w}, p\right) \leq g^{\mathrm{wd}}\left(\mathbf{x}^k \middle| \mathbf{w}, p\right) \end{aligned} \tag{6}$$

实际：

PF未知 $\Longrightarrow$ z*未知

转换思路：

假设要找的聚合函数为F1，F1求得的当前最优解为x1。所以问题转变为找最优解为x1的聚合函数。而X1的特点是兼备了收敛性和分布性。

目标2：找出能够得到离权重向量最近的最优解的聚合函数

解决方案：PaS

Algorithm 1: Pareto Adaptive L_p Scalarizing Approximation, PaS

Input: current population S, current generation index: gen, maximum generation index: $maxGen$, weight vector $\mathbf{w}$, candidate p set: $P = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \infty\}$

Output: p value for subproblem with weight vector $\mathbf{w}$

1 **if** $rand \geq gen/maxGen$ **then**
2 **foreach** $p^k \in P$ **do**
3 $\mathbf{z}^k \leftarrow \arg\min_{\mathbf{x} \in S} \left(g^{wd}(\mathbf{x}|\mathbf{w}, p^k)\right)$;
4 **end**
5 $\mathbf{z} \leftarrow \arg\min_{\mathbf{z}^k} \mathtt{dist}\left(\mathbf{z}^k, \mathbf{w}\right)$;
6 Set p to be the smallest p^k with $\mathbf{z} = \mathbf{z}^k$;
7 **end**

PaS举例

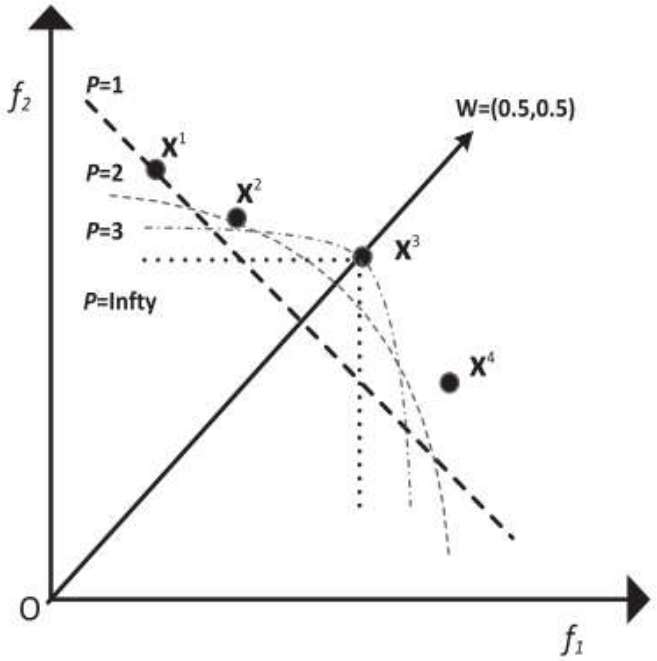


Fig. 11. Simple bi-objective minimization example for illustrating the PaS procedure.

TABLE I
SCALAR VALUES OF SOLUTIONS USING VARIOUS L_p SCALARIZING METHODS

$\mathbf{w} = (0.5, 0.5)$	$p = 1$	$p = 2$	$p = 3$	$p = \infty$	distance
$f(\mathbf{x}^1) = (0.2, 0.7)$	0.45	0.3640	0.3506	0.35	0.3536
$f(\mathbf{x}^2) = (0.3, 0.65)$	0.475	0.3579	0.3286	0.325	0.2475
$f(\mathbf{x}^3) = (0.55, 0.55)$	0.55	0.3889	0.3270	0.275	0
$f(\mathbf{x}^4) = (0.8, 0.4)$	0.6	0.4472	0.4061	0.40	0.2828

假设当前有4个解，x1,x2,x3,x4，权重向量为（0.5,0.5）。
当p=1时，最优解为x1
当p=2时，最优解为x2
当p=3时，最优解为x3
当p=inf时，最优解为x3
在x1,x2,x3中，x3离权重向量的距离(d2) 最小。
选择x3的聚合函数有，p=3和p=inf，所以根据公式，要在符合约束条件的p中选择值最小的p。
所以该例子中，**聚合函数p值更新为3。**

P集合的选定

从理论上说，p∈P={1,2,3,…, ∞}
但是，遍历所有P中所有p值来确定聚合函数
显然不可能

所以将P修改为有限集P={1,2,3,…,10, ∞}

这样做的依据是：

- p=10与p= ∞的等高线已经很接近。
- 且右图实验表明， 搜索能力从p=10到p= ∞
  并没有大幅度下降

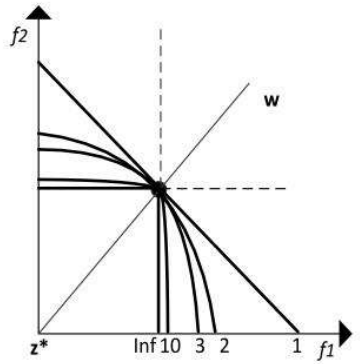


Fig. 3. Contour curves of the L_p scalarizing method with different p values.

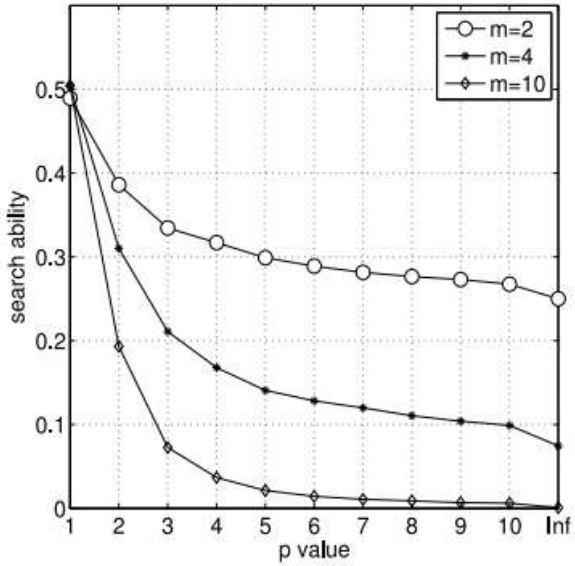


Fig. 15. Change of search ability of different L_p scalarizing methods for two-, four-, and ten-objective problems.

PaS可行的前提：

The best solution identified by various *g*wd*(***x**|**w**, *pk)* is different with large probability

实验证明该前提：

ten nondominated solutions are generated in the two-objective space and nine evenly distributed
weight vectors are considered, which are **w**1 = *(*0.1, 0.9*)*,**w**2 = *(*0.2, 0.8*)*,..., **w**9 = *(*0.9, 0.1*)*. We then find the best solution, **z***k* for each *g*wd*(***x**|**w**, *pk)*.
The experiment is repeated 100 times

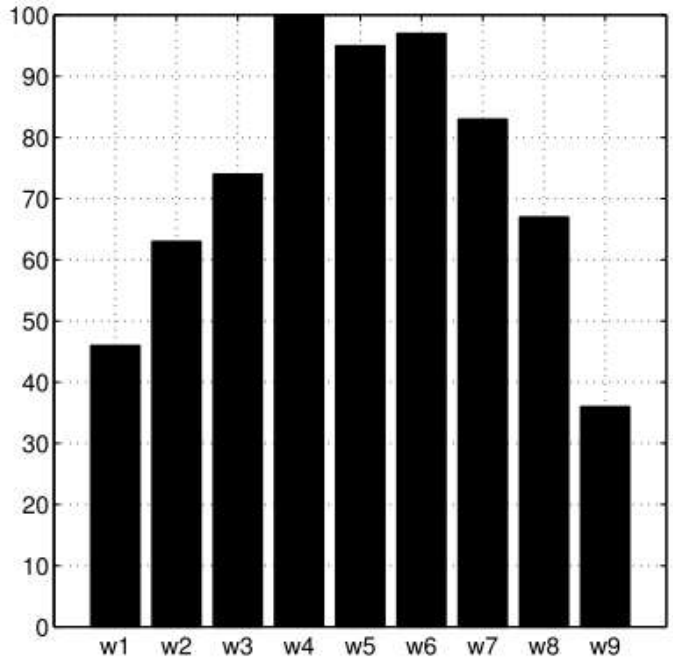


Fig. 10. Number of times (*y*-axis) that the best solutions are different for different weight vectors (*x*-axis).

MOEA/D-PaS

Algorithm 2: MOEA/D Using Pareto Adaptive L_p Scalarizing Methods, MOEA/D-PaS

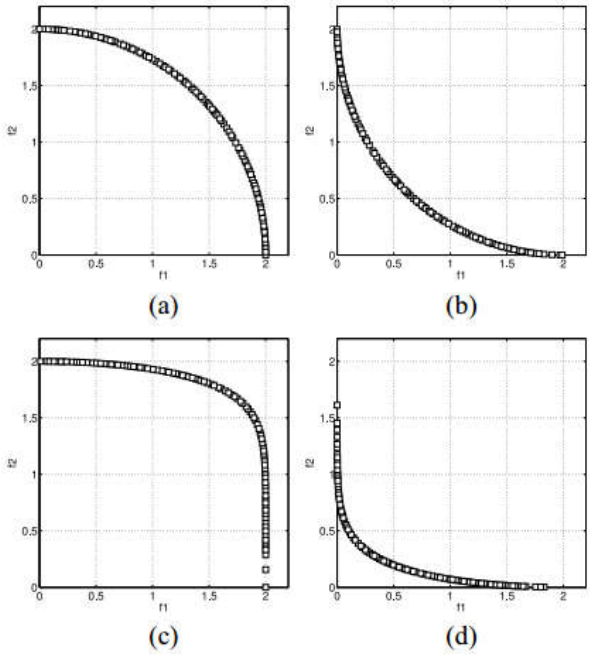
Input: initial population, $S \leftarrow \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^N\}$, initial weights, $W \leftarrow \{\mathbf{w}^1, \mathbf{w}^2, \ldots, \mathbf{w}^N\}$, selection neighbourhood size, T, replacement neighbourhood size, nr, maximum generation index, *maxGen*

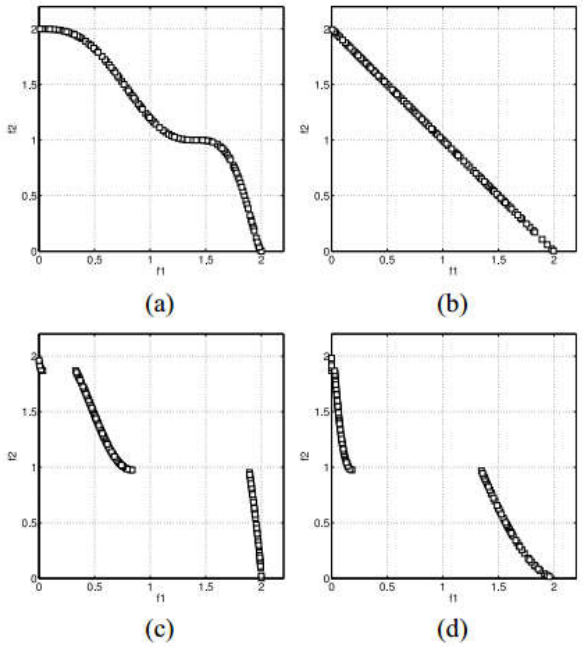
Output: S

Initialize the candidate p set, $P = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \infty\}$;
Evaluate the objective function values of the initial S;
Update *ideal* and *nadir* vectors, $\mathbf{z}^*$ and $\mathbf{z}^{nad}$;
Randomly assign each weight, $\mathbf{w}^i$ with a candidate solution, $\mathbf{x}^i$;
Calculate the Euclidean distance between weights, $\mathbf{w}^i$ and $\mathbf{w}^j$, $i, j \in 1, 2, \ldots, N$;
Find T neighboring weights $B(\mathbf{w}^i)$ of $\mathbf{w}^i$ and identify the associated neighboring solutions $B(\mathbf{x}^i)$ of $\mathbf{x}^i$;
Set $p^i = \infty, i = 1, 2, \ldots, N$, $gen = 0$, the mating pool $Q = \emptyset$, the probability of mating restriction $\delta = 0.9$;
while $gen \leq maxGen$ **do**
  for $i \leftarrow 1$ **to** N **do**
    if $rand < \delta$ **then**
      $Q \leftarrow B(\mathbf{x}^i)$;
    else
      $Q \leftarrow S$;
    end
    Randomly select three solutions $\mathbf{x}^{r1}$, $\mathbf{x}^{r2}$ and $\mathbf{x}^{r3}$ from the mating pool, Q;
    Generate a new solution $\mathbf{x}^{new}$ by performing differential evolution (DE) and polynomial mutation (PM) operators;
    Evaluate the objective value of $\mathbf{x}^{new}$, and update $\mathbf{z}^*$ and $\mathbf{z}^{nad}$;
    $c \leftarrow 0$;
    while $c \leq nr || Q \neq \emptyset$ **do**
      randomly pick an index j from Q;
      if $g^{wd}(\mathbf{x}^{new}|\mathbf{w}^i, p^i) < g^{wd}(\mathbf{x}^j|\mathbf{w}^i, p^i)$ **then**
        $\mathbf{x}^j \leftarrow \mathbf{x}^{new}$;
        $c \leftarrow c + 1$
      end
      remove $\mathbf{x}^j$ from Q;
    end
  end
  Update the p^i, $i = 1, 2, \ldots, N$ using Algorithm 1;
  $gen \leftarrow gen + 1$;
end

实验

- 测试集WFG4X
- 涵盖各种形状的PF





对比算法

- MOEA/D
- MOEA/D-AS

  MOEA/D with an adaptive use of the weighted sum and Chebyshev methods),
- MOEA/D-SS

  MOEA/D with a simultaneous use of the weighted sum and Chebyshev methods.

实验结果

HV值：
代表分布性，越大越好

	MOEA/D-PaS	MOEA/D	MOEA/D-AS	MOEA/D-SS
2-objective problems				
WFG41-2	0.3138(0.0031)	0.2982(0.0052)$^-$	0.2876(0.0019)$^-$	0.2965(0.0052)$^-$
WFG42-2	0.7887(0.0038)	0.7710(0.0047)$^-$	0.7784(0.0034)$^-$	0.7790(0.0025)$^-$
WFG43-2	0.1907(0.0056)	0.1797(0.0175)$^-$	0.1796(0.0047)$^-$	0.1839(0.0013)$^-$
WFG44-2	0.9051(0.0038)	0.8922(0.0050)$^-$	0.9083(0.0047)$^=$	0.9037(0.0022)$^=$
WFG45-2	0.4291(0.0029)	0.4145(0.0072)$^-$	0.4076(0.0041)$^-$	0.4102(0.0019)$^-$
WFG46-2	0.5460(0.0021)	0.5286(0.0071)$^-$	0.5307(0.0013)$^-$	0.5392(0.0013)$^-$
WFG47-2	0.4625(0.0123)	0.4323(0.0197)$^-$	0.4429(0.0192)$^-$	0.4703(0.0028)$^=$
WFG48-2	0.5981(0.0232)	0.5094(0.0052)$^-$	0.4999(0.0024)$^-$	0.5398(0.0277)$^-$
4-objective problems				
WFG41-4	0.5861(0.0028)	0.4668(0.0097)$^-$	0.4876(0.0084)$^-$	0.5516(0.0045)$^-$
WFG42-4	0.9100(0.0079)	0.8828(0.0072)$^-$	0.8958(0.0142)$^=$	0.9015(0.0090)$^-$
WFG43-4	0.3617(0.0443)	0.3354(0.0052)$^-$	0.3575(0.0047)$^-$	0.3619(0.0020)$^=$
WFG44-4	0.9424(0.0053)	0.9401(0.0048)$^=$	0.9440(0.0111)$^=$	0.9491(0.0039)$^=$
WFG45-4	0.6609(0.0045)	0.5614(0.0151)$^-$	0.5951(0.0099)$^-$	0.6455(0.0117)$^-$
WFG46-4	0.8147(0.0194)	0.7318(0.0218)$^-$	0.7731(0.0264)$^-$	0.8255(0.0030)$^+$
WFG47-4	0.6911(0.0049)	0.4855(0.0582)$^-$	0.6441(0.0157)$^-$	0.6789(0.0075)$^-$
WFG48-4	0.6956(0.0120)	0.6564(0.0258)$^-$	0.7240(0.0206)$^+$	0.7071(0.0207)$^=$
7-objective problems				
WFG41-7	0.7204(0.0340)	0.5319(0.0212)$^-$	0.6286(0.0452)$^-$	0.6149(0.0617)$^-$
WFG42-7	0.9556(0.0011)	0.9500(0.0069)$^=$	0.9448(0.0009)$^=$	0.9438(0.0052)$^=$
WFG43-7	0.4252(0.0219)	0.3375(0.0413)$^-$	0.3667(0.0501)$^-$	0.3319(0.0509)$^-$
WFG44-7	0.9487(0.0094)	0.9341(0.0027)$^-$	0.9454(0.0095)$^=$	0.9519(0.0104)$^=$
WFG45-7	0.7596(0.0122)	0.6293(0.0035)$^-$	0.6936(0.0054)$^-$	0.6527(0.0005)$^-$
WFG46-7	0.8884(0.0084)	0.8112(0.0165)$^-$	0.8385(0.0379)$^-$	0.8399(0.0027)$^-$
WFG47-7	0.6584(0.0230)	0.5011(0.0191)$^-$	0.5222(0.0422)$^-$	0.5607(0.0272)$^-$
WFG48-7	0.7497(0.0101)	0.6941(0.0240)$^-$	0.7654(0.0138)$^=$	0.7447(0.0283)$^=$
♯. $-$ / $=$ /+		22/2/0	17/6/1	15/8/1

实验结果

C值：
代表收敛性，越大越好

【提示】
C(A,B) > C(B,A)，
说明A的收敛程度比B好

	C(MOEA/D,PaS)		C(PaS,MOEA/D)	C(AS,PaS)		C(PaS,AS)	C(SS,PaS)		C(PaS,SS)
2-objective problems									
WFG41	0.0052(0.0104)	<	0.9640(0.0426)	0(0)	<	0.9748(0.0170)	0.0401(0.0463)	<	0.6904(0.0802)
WFG42	0.0984(0.0668)	<	0.7846(0.0464)	0.0346(0.0691)	<	0.7259(0.0797)	0.1481(0.0574)	<	0.5528(0.1256)
WFG43	0.1691(0.0867)	<	0.4279(0.0481)	0(0)	<	0.8606(0.0627)	0.1825(0.0861)	<	0.4833(0.0927)
WFG44	0.0026(0.0052)	<	0.9384(0.0474)	0.0524(0.0835)	<	0.5684(0.0774)	0.3777(0.0462)	=	0.3894(0.1057)
WFG45	0.0111(0.0222)	<	0.7765(0.0330)	0.0082(0.0163)	<	0.8107(0.0826)	0.0078(0.0156)	<	0.7981(0.1276)
WFG46	0.0438(0.0876)	<	0.7548(0.0404)	0.0025(0.0050)	<	0.7662(0.0732)	0.1065(0.0886)	<	0.3749(0.1032)
WFG47	0.0357(0.0714)	<	0.6808(0.0327)	0.0073(0.0085)	<	0.6080(0.0837)	0.0299(0.0294)	<	0.6310(0.1941)
WFG48	0.0740(0.0921)	<	0.5952(0.0458)	0.0664(0.0668)	<	0.4243(0.0583)	0.1164(0.0225)	=	0.1027(0.1011)
4-objective problems									
WFG41	0.0020(0.0041)	<	0.3495(0.0427)	0(0)	<	0.4768(0.0337)	0.0446(0.0143)	<	0.0580(0.0341)
WFG42	0(0)	<	0.1355(0.0642)	0.0023(0.0045)	<	0.1706(0.0864)	0.0045(0.0052)	<	0.0090(0.0027)
WFG43	0(0)	<	0.5230(0.0870)	0(0)	<	0.3971(0.0948)	0.0227(0.0344)	<	0.2243(0.0964)
WFG44	0.2250(0.1039)	<	0.1153(0.0894)	0.0073(0.0092)	<	0.2385(0.1002)	0.0224(0.0122)	<	0.0580(0.0882)
WFG45	0.0021(0.0042)	<	0.2781(0.0276)	0(0)	<	0.3180(0.0605)	0.0301(0.0156)	=	0.0497(0.0227)
WFG46	0.0044(0.0051)	<	0.1118(0.0638)	0.0044(0.0251)	<	0.1467(0.0686)	0.0173(0.0097)	=	0.0190(0.0113)
WFG47	0.0024(0.0047)	<	0.2629(0.0795)	0.0194(0.0243)	<	0.1298(0.0461)	0.0474(0.0246)	<	0.1418(0.0265)
WFG48	0.0208(0.0417)	<	0.4610(0.0898)	0.1725(0.0928)	=	0.1395(0.0819)	0.0612(0.0315)	=	0.0636(0.0922)
7-objective problems									
WFG41	0(0)	<	0.0633(0.0368)	0.0143(0.0280)	<	0.0667(0.0278)	0.0186(0.0148)	=	0.0169(0.0120)
WFG42	0.0263(0.0137)	<	0.0551(0.0178)	0.0065(0.0450)	<	0.0809(0.0427)	0.0261(0.0001)	<	0.0840(0.0369)
WFG43	0.0156(0.0321)	<	0.2500(0.0436)	0.0732(0.0698)	<	0.2198(0.0911)	0(0)	<	0.2708(0.0283)
WFG44	0(0)	<	0.2692(0.0503)	0.0106(0.0469)	<	0.0532(0.0376)	0.0288(0.0180)	<	0.1892(0.0138)
WFG45	0(0)	<	0.2386(0.1076)	0(0)	<	0.1379(0.0607)	0(0)	<	0.1538(0.0709)
WFG46	0.0797(0.0239)	<	0(0)	0.0162(0.0045)	=	0.0126(0.0089)	0.0155(0.0313)	=	0.0102(0.0304)
WFG47	0.0800(0.0286)	<	0.2027(0.1093)	0.0120(0.0551)	<	0.2500(0.0029)	0.0482(0.0225)	<	0.4821(0.2415)
WFG48	0(0)	<	0.2020(0.1185)	0(0)	<	0.7419(0.5246)	0(0)	<	0.4211(0.2270)
	♯ </=/>	24/0/0		♯ </=/>	22/2/0		♯ </=/>	17/7/0	

P值的变化

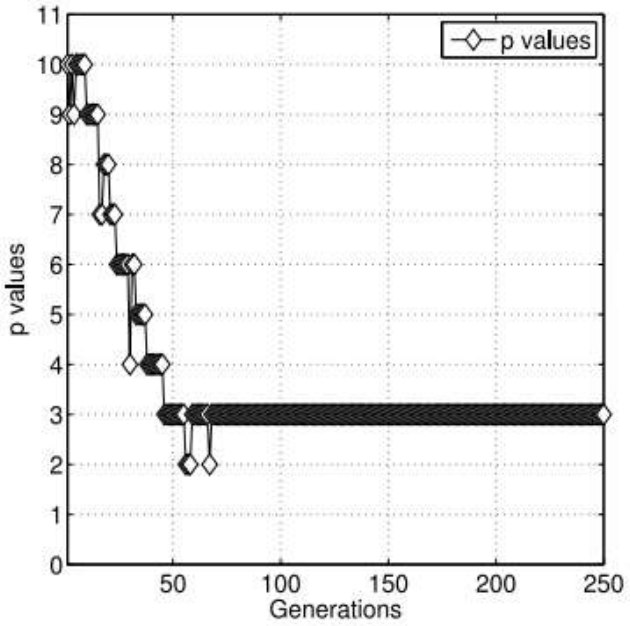


Fig. 14. Obtained p values for WFG41-4 over generations.

在数据集WFG41-4上，搜索方向w=(0.25,0.25,0.25,0.25)对应的p值随迭代的变化

P收敛为3，所以用p=3的聚合函数便能够找出所有最优解

理想效果

TABLE IV
HV METRIC COMPARISON RESULTS (MEAN AND STD) OF MOEA/D-PAS, MOEA/D-PAS2, AND MOEA/D-PFT FOR THE WFG41 PROBLEMS. THE BEST RESULTS ARE MARKED IN BOLD TYPE

	MOEA/D-PaS	MOEA/D-PaS2	MOEA/D-Pft
WFG41-2	0.3138(0.0031)	0.3013(0.0051)	**0.3234(0.0018)**
WFG41-4	0.5861(0.0028)	0.5445(0.0078)	**0.6094(0.0019)**
WFG41-7	0.7204(0.0340)	0.6903(0.0443)	**0.7504(0.0110)**

	C(PaS,PaS2)	>	C(PaS2,PaS)	C(PaS,Pft)	<	C(Pft,PaS)
WFG41-2	0.61(0.0121)	<	0.013(0.0024)	0.0016(0.0002)	>	0.0541(0.0122)
WFG41-4	0.47(0.0098)	<	0.04(0.0257)	0.0056(0.0003)	>	0.1005(0.0454)
WFG41-7	0.39(0.0098)	<	0.11(0.0317)	0.0316(0.0049)	>	0.1804(0.0315)

【提示】 C(A,B) > C(B,A)，说明A的收敛程度比B好

MOEA/D-Pft：在预知PF形状的情况下设定p值
MOEA/D-PaS：文章提出
MOEA/D-PaS2：更新p值时，没有使用if条件--rand≥gen/maxGen

HV:

MOEA/D-Pft > MOEA/D-PaS

C:

MOEA/D-Pft > MOEA/D-PaS
MOEA/D-PaS > MOEA/D-PaS2

【结论】

1. 自适应对收敛性产生了一定的负作用
2. 加上条件，rand≥gen/maxGen，减轻了该负作用

谢谢！