Montreal Optimization Days May 2004

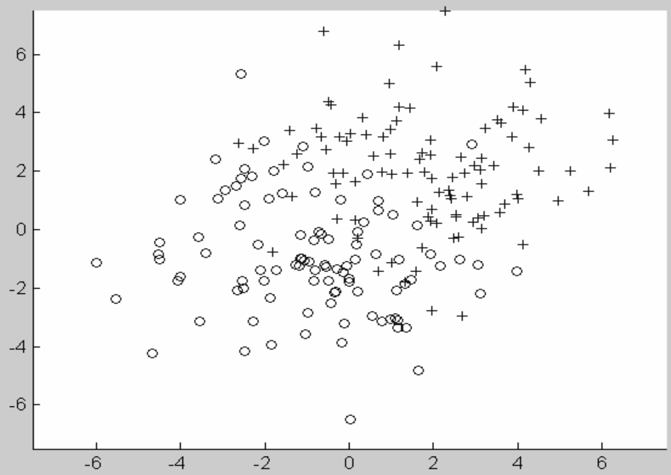
Arbitrary-norm Separation by Variable Neighborhood Search

Pierre Hansen (pierre.hansen@gerad.ca)
Alejandro Karam (alejandro.karam@hec.ca)

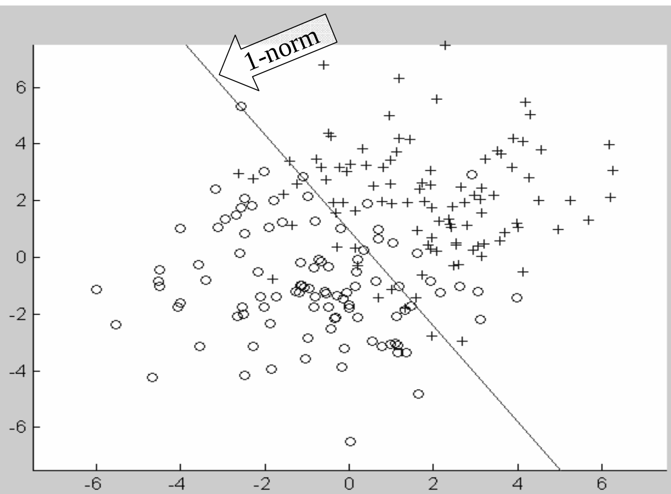
Today's Talk:

- The Problem: L_p-Norm Separation
- Solution Approach
- Variable Neighborhood Search
- Numerical Tests
- Conclusions

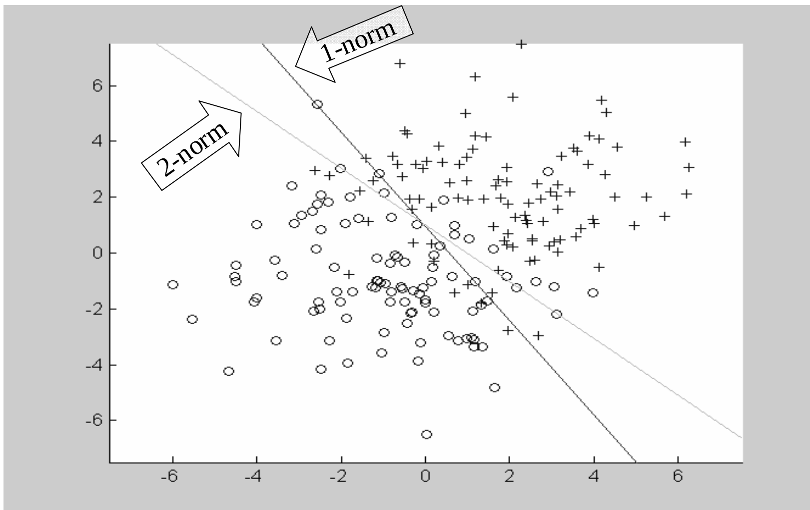
L_p-Norm Separation



L_p-Norm Separation



L_p-Norm Separation



L_p-Norm Separation

The distance from a point $z \in \mathbb{R}^n$ to a plane $P = \{x \,|w'x = \gamma\}$ in norm L_p is

$$\frac{|w'x - \gamma|}{\|w\|_p'} \tag{1}$$

where $\|w\|_p'$ is the dual norm of L_p.

For $1 < p < \infty$, $\|w\|_p' = \|w\|_q$ where q is such that $\frac{1}{p} + \frac{1}{q} = 1$.

We therefore compute it as

$$\|w\|_p' = \sqrt[\frac{p-1}{p}]{\sum w^{\frac{p}{p-1}}} \tag{2}$$

L_p-Norm Separation

- The problem is

$$\min_{(w \in \mathbb{R}^n, \gamma \in \mathbb{R})} \left\{ \frac{\sum_{i=1}^{m} \max\{-w'A_i+\gamma,0\} + \sum_{j=1}^{k} \max\{w'B_j-\gamma,0\}}{\|w\|_p'} \right\}$$

- where $\|w\|_p' = \sqrt[\frac{p-1}{p}]{\sum w^{\frac{p}{p-1}}}$

UGLY

L_p-Norm Separation

- The problem is

$$\min_{(w \in \mathbb{R}^n, \gamma \in \mathbb{R})} \left\{ \frac{\sum_{i=1}^{m} \max\{-w'A_i+\gamma,0\} + \sum_{j=1}^{k} \max\{w'B_j-\gamma,0\}}{\|w\|_p'} \right\}$$

- where $\|w\|_p' = \sqrt[\frac{p-1}{p}]{\sum w^{\frac{p}{p-1}}} = 1$

- Linearize the *max* {•} operator and impose constraint on denominator
- Exact solution for norms 1, 2 and ∞ for moderately sized instances

Today's Talk:

- The Problem: L_p-Norm Separation
- Solution Approach
- Variable Neighborhood Search
- Numerical Tests
- Conclusions

Solution Approach

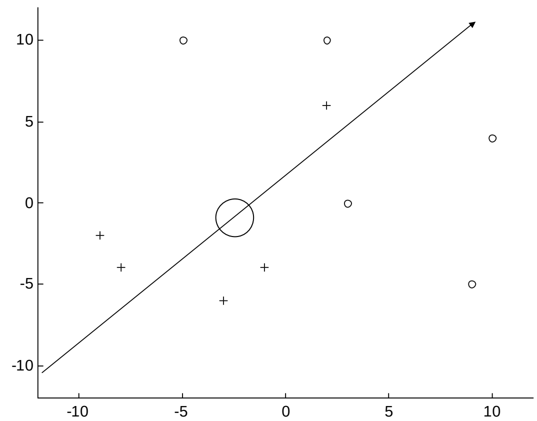
- Decompose the search of the gradient *w* from that of the offset to the origin *γ*
- Note that for fixed *w* , conversion from Euclidean norm is a constant rescaling

$$P = \{x \,|w'x = \gamma\}$$

$$\frac{|w'x - \gamma|}{\|w\|_p'}$$

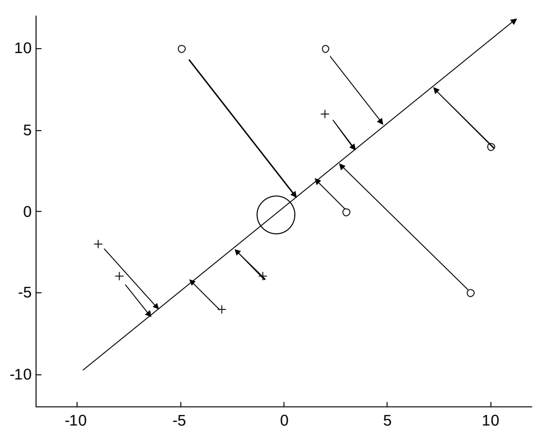
Solution Approach

The decomposition:
• fix a direction
• project points into ray
• find best position for plane



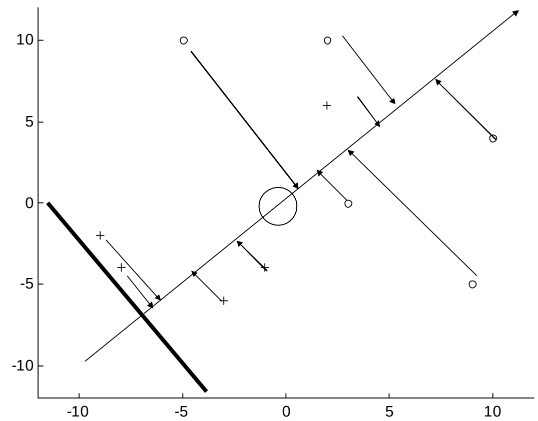
Solution Approach

The decomposition:
• fix a direction
• project points into ray
• find best position for plane



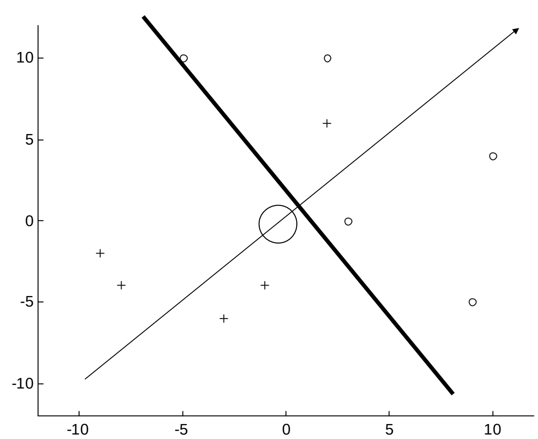
Solution Approach

The decomposition:
- fix a direction
- project points into ray
- find best position for plane



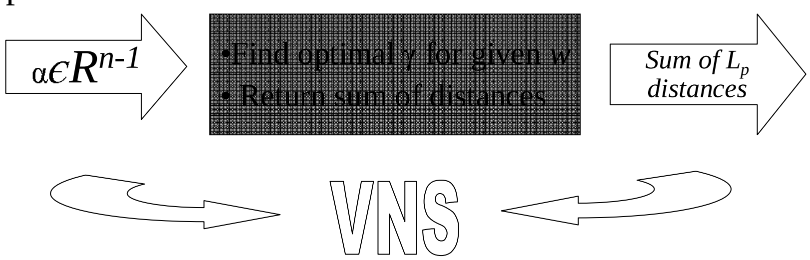
Solution Approach

The decomposition:
- fix a direction
- project points into ray
- find best position for plane



Solution Approach

- We characterize the directions by n-1 angles $\boldsymbol{\alpha}=(\alpha_1, \alpha_2, \ldots, \alpha_{n-1})$
- Oracle F($\boldsymbol{\alpha}$) that takes a direction $\boldsymbol{\alpha}$, finds the optimal offset **in that direction** and returns the corresponding sum of distances of misclassified points



Today's Talk:

- The Problem: L_p-Norm Separation
- Solution Approach
- Variable Neighborhood Search
- Numerical Tests
- Conclusions

Variable Neighborhood Search

Key ideas:

– Exploit valuable information in local minima

– Random perturbations increasingly "far" from incumbent solution

Ingredients:

– Local descent method

– Distance on solution space

} problem dependent

Variable Neighborhood Search

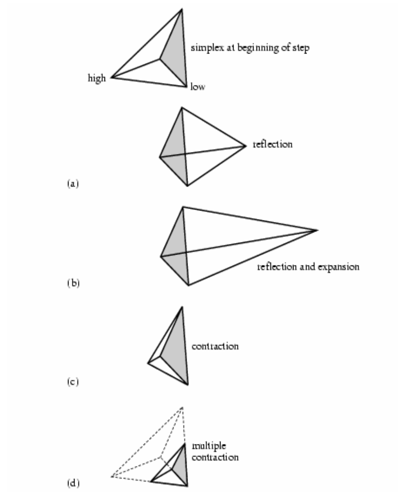
- Select starting solution *x*
- Define neighborhood structure $N_k(x)$ with k=1,…,*kmax*
- Define stopping criteria

```
WHILE [stop criteria not met]
    k←1
    WHILE [k < kmax]
        ▪ generate x'ε N_k (x)
        ▪ start local search at x' to find x''
        ▪ IF [x'' is better]
            THEN
                recenter x← x''
                and reset k←1
            ELSE
                try next neighborhood:
                k←k+1
          ENDIF
    ENDWHILE
ENDWHILE
```

VNS for L_p-norm separation

the local descent

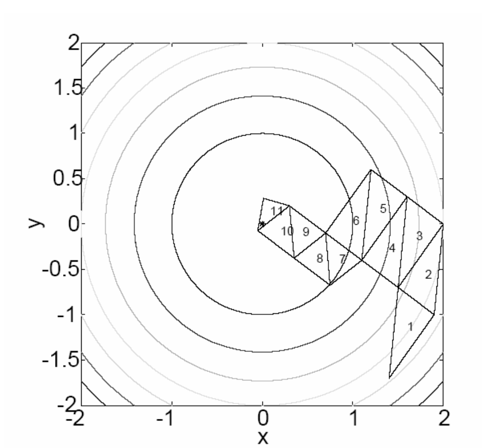
- Downhill simplex method of (Nelder and Mead 1965)
- Most basic implementation from Numerical Recipes in C



VNS for L_p-norm separation

the local descent

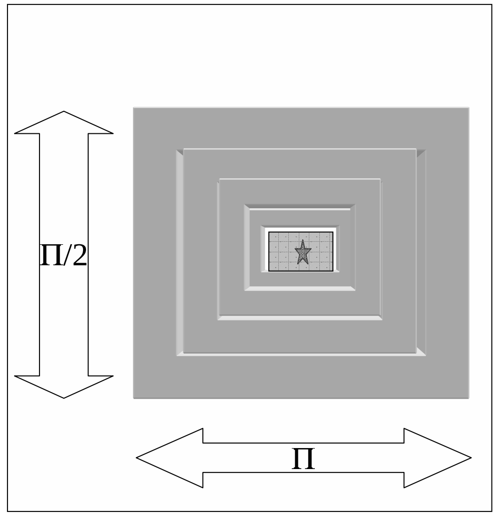
- Downhill simplex method of (Nelder and Mead 1965)
- Most basic implementation from Numerical Recipes in C



VNS for L_p-norm separation

Neighborhood structure

- Outer shells of nested rectangular hyper-boxes
- Outmost box measures Π/2 in n-2 directions and Π in the last



Today's Talk:

- The Problem: L_p-Norm Separation
- Solution Approach
- Variable Neighborhood Search
- Numerical Tests
- Conclusions

Numerical Tests

- Benchmarks to exact solutions
- Acceleration of exact methods

} Random problems

- Effect of norm on real data

} UCI repository

Numerical Tests

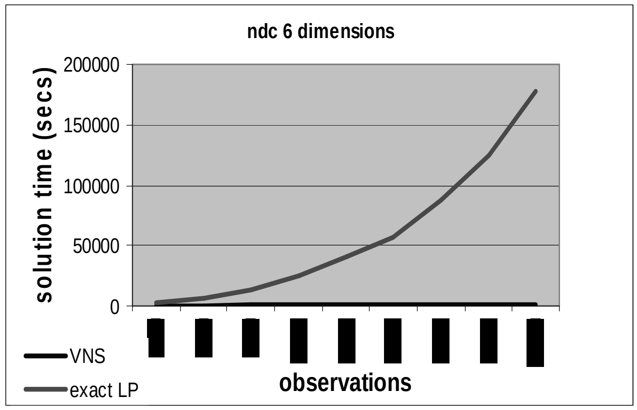
Benchmarks to exact solutions

problem	dims	obs	error	time exact	time VNS
ndc6d4k25v	6	4000	0	2505	418
ndc6d6k25v	6	6000	0	6554	553
ndc6d8k25v	6	8000	0	12737	810
ndc6d10k25v	6	10000	0	25600	1144
ndc6d12k25v	6	12000	0	41088	681
ndc6d14k25v	6	14000	0	57513	1040
ndc6d16k25v	6	16000	0	86845	860
ndc6d18k25v	6	18000	0	124152	936
ndc6d20k25v	6	20000	0	177798	1649

Numerical Tests

Benchmarks to exact solutions

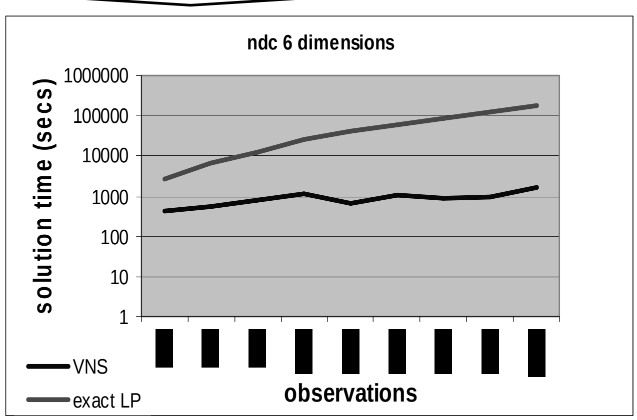
- 100% accuracy on this set
- exact solution time explodes



Numerical Tests

Benchmarks to exact solutions

- 100% accuracy on this set
- exact solution time explodes



Numerical Tests

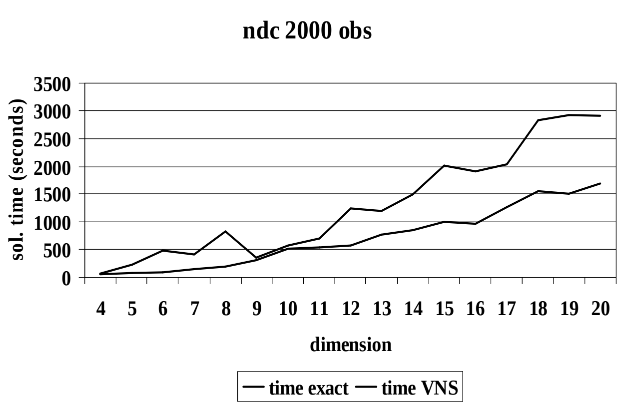
Benchmarks to exact solutions

problem	dims	obs	error	time exact	time VNS
ndc4d2k25v	4	2000	0%	61.909	53.417
ndc5d2k25v	5	2000	0%	228.628	71.883
ndc6d2k25v	6	2000	0%	482.083	83.3
ndc7d2k25v	7	2000	0%	420.064	154.082
ndc8d2k25v	8	2000	0%	829.883	187.53
ndc9d2k25v	9	2000	17%	351.696	315.153
ndc10d2k25v	10	2000	0%	579.513	511.406
ndc11d2k25v	11	2000	11%	699.115	535.63
ndc12d2k25v	12	2000	22%	1240.584	580.004
ndc13d2k25v	13	2000	0%	1189.791	762.987
ndc14d2k25v	14	2000	0%	1497.804	857.634
ndc15d2k25v	15	2000	0%	2006.214	1000.119
ndc16d2k25v	16	2000	26%	1906.982	969.023
ndc17d2k25v	17	2000	5%	2036.017	1272.63
ndc18d2k25v	18	2000	13%	2833.094	1553.303
ndc19d2k25v	19	2000	7%	2916.974	1507.988
ndc20d2k25v	20	2000	0%	2905.418	1699.233

- The method starts to have a hard time in dimensions above about 9
- On easier cases, reasonable results in higher dimensions
- Time advantage less dramatic as dimension grows

Numerical Tests

Benchmarks to exact solutions

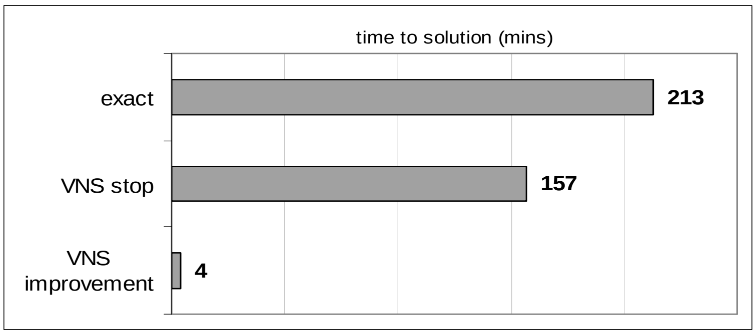


- The method starts to have a hard time in dimensions above about 9
- On easier cases, reasonable results in higher dimensions
- Time advantage less dramatic as dimension grows

Numerical Tests

Benchmarks to exact solutions

A larger example: ndc 8 dim., 100,000 obs., 1-norm solution



Numerical Tests

Benchmarks to exact solutions

Cases from UCI Repository:

problem	dims	obs	n1 error	n2 error	ninf error
vowels20k	16	20000	381.48%	N/A	N/A
glass_windows	9	214	0.47%	0.66%	0.87%
echocardiogram	6	62	0.00%	0.00%	0.00%
diabetes	8	768	0.00%	0.00%	0.00%
housing	13	506	333.42%	0.06%	0.70%
ionosphere_noAM	32	351	7.05%	N/A	N/A
hepatitis_noEDC	9	112	1.96%	N/A	0.24%
breast10dPCA	10	569	0.23%	N/A	0.02%

Numerical Tests

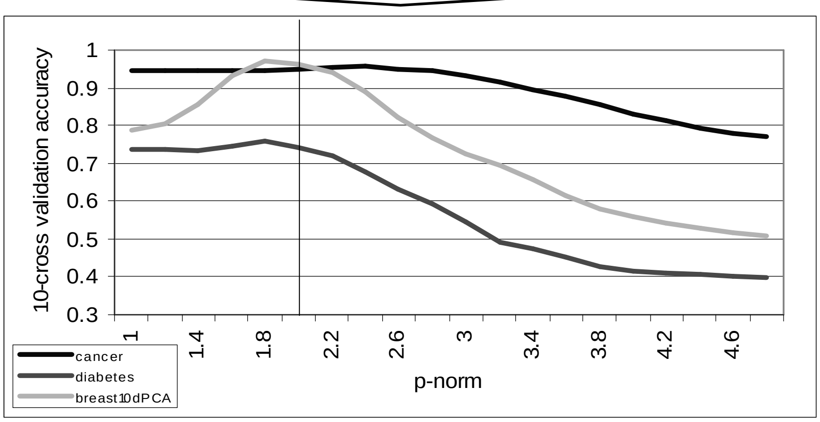
Acceleration of exact methods

norm inf (MIP)				exact	exact with VNS bound	VNS improv-ment	VNS stop time
problem	dim	obs	error	hours	hours	hours	hours
test5d100k	5	100,000	1.6E-08	92.92	25.93	0.13	0.58
test7d100k	7	100,000	3.8E-08	282.27	87.46	1.10	1.24
test10d100k	10	100,000	1.5E-05	211.16	5.41	0.64	5.28

- MIP solved with CPLEX
- Slightly less dramatic results for 1-norm, much more impressive for many 2-norm cases

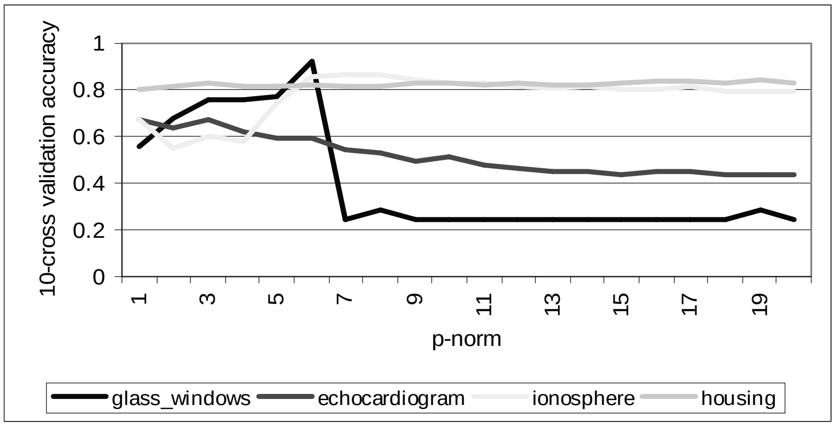
Numerical Tests

Effect of norm on real data



Numerical Tests

Effect of norm on real data



Today's Talk:

- The Problem: L_p-Norm Separation
- Solution Approach
- Variable Neighborhood Search
- Numerical Tests
- Conclusions

Conclusions

- It seems to work!
- But it can be improved: this is a first approximation
- Reasonable performance on moderate dimensions

Conclusions

- Method scales up very well: important in Data Mining
- Remarkable acceleration of exact methods

Conclusions

- Interesting results on real life data with varying norm:
  - Impossible to obtain otherwise (i.e. exactly)
  - Interpretation is still an open challenge