Introduction to Analyses of Adaptive Stochastic Search Methods for Global Optimization

Zelda B. Zabinsky

Industrial Engineering Program
University of Washington, Seattle, WA

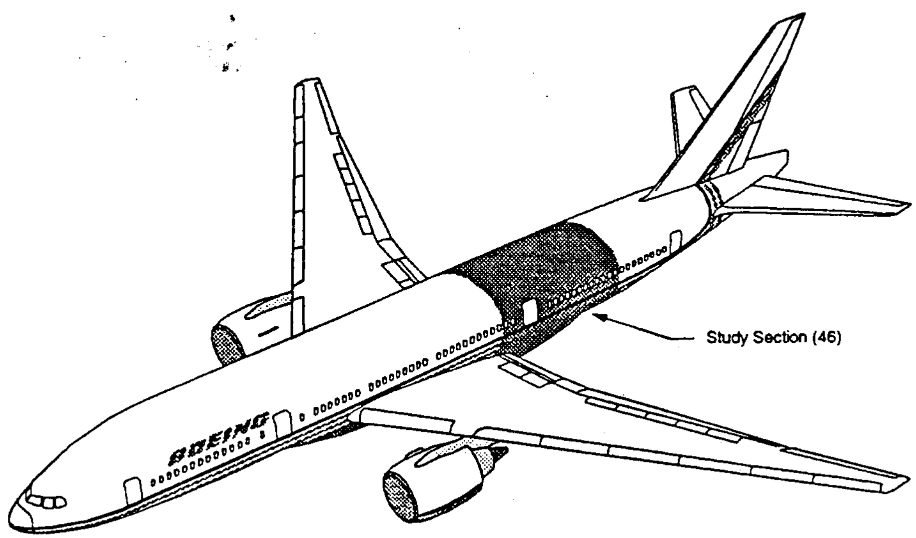
September 2001

Overview

- Practical global optimization problems in engineering design
- Theoretical performance of stochastic adaptive search methods
- Algorithms based on Hit-and-Run to approximate theoretical performance
- Engineering design problems and manufacturing tolerances

Problems in Engineering Design

- Need to consider manufacturing and cost considerations early in the design process, because a large percentage of cost is locked in at preliminary design
- Use optimization in preliminary design to quantify tradeoffs

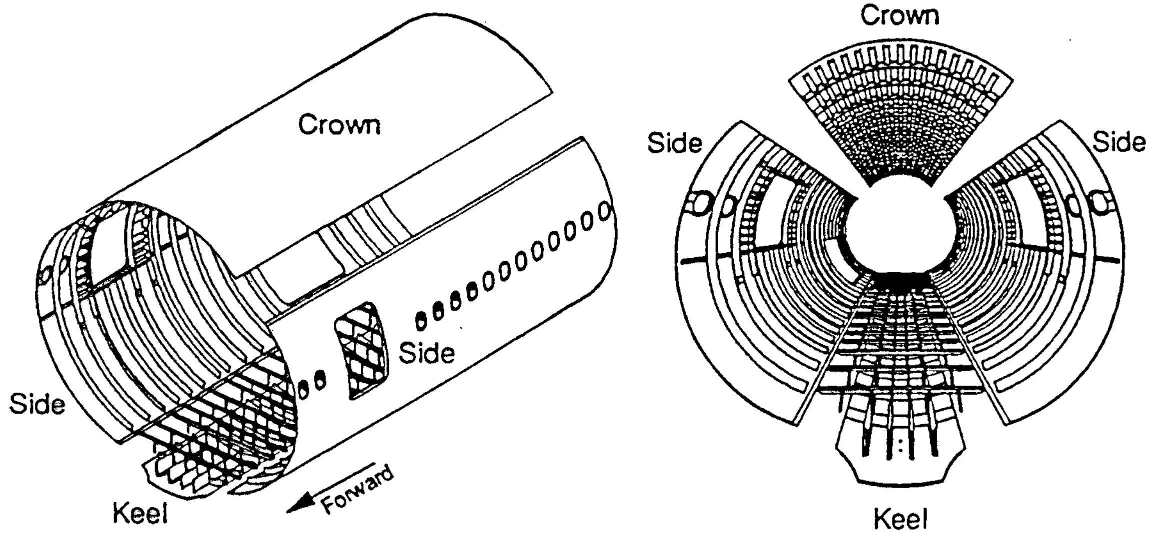


[NASA Contractor Report 4732, April 1997]

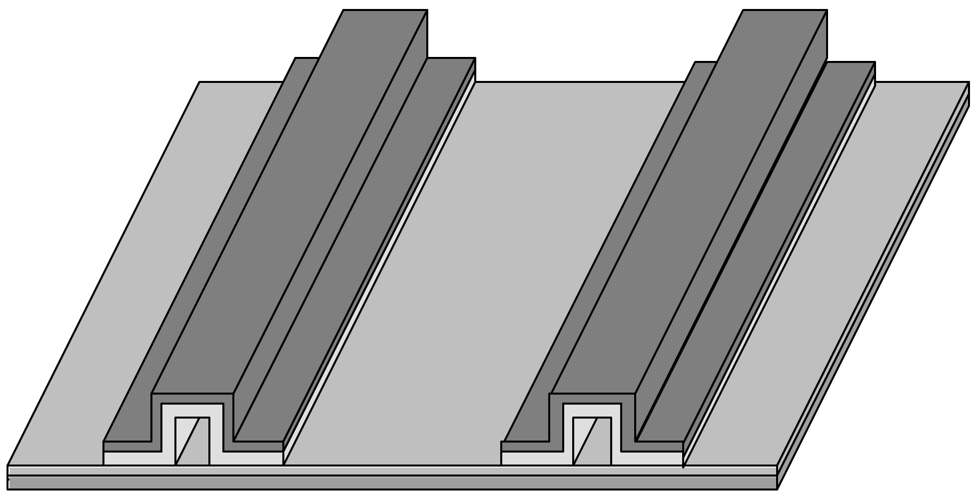
Composite Structures

- Composite laminates
  - fibers in a resin, plies bonded together
  - e.g. graphite epoxy
- Attractive for light weight structures
  - high strength- and stiffness-to-weight ratios
  - can design the *material* as well as the structure

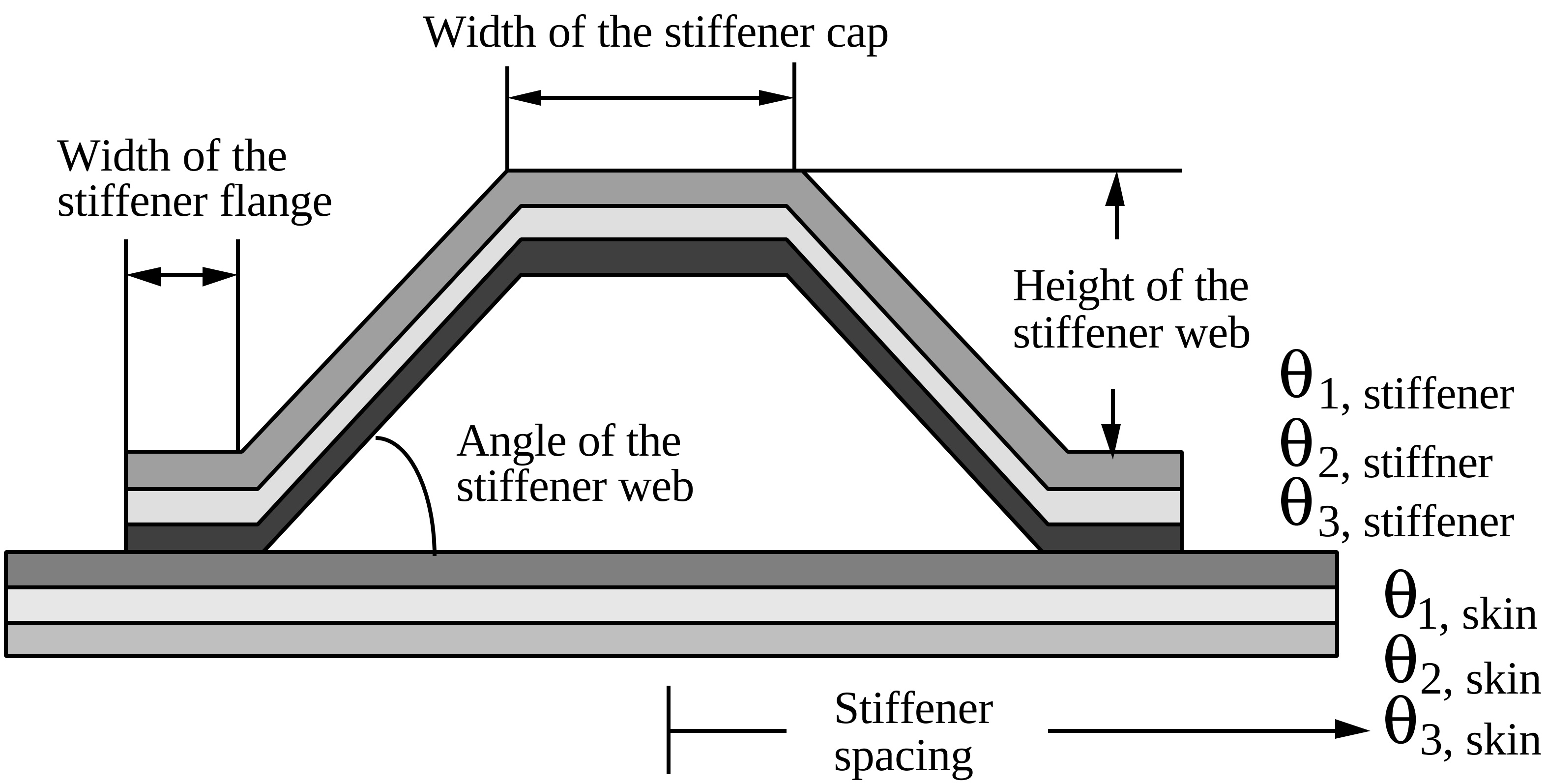
Aircraft Panels



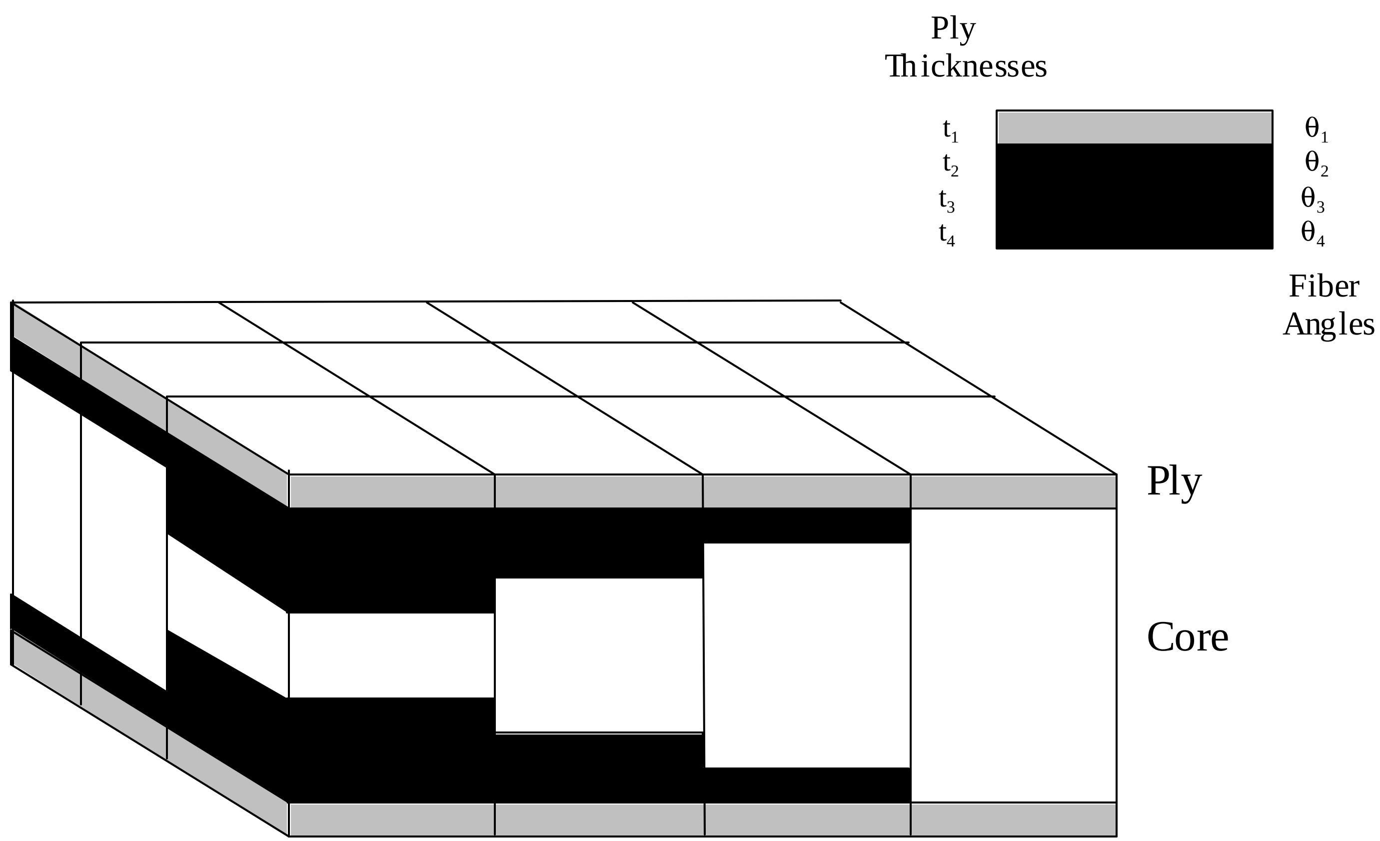
Hat Stiffened Panel



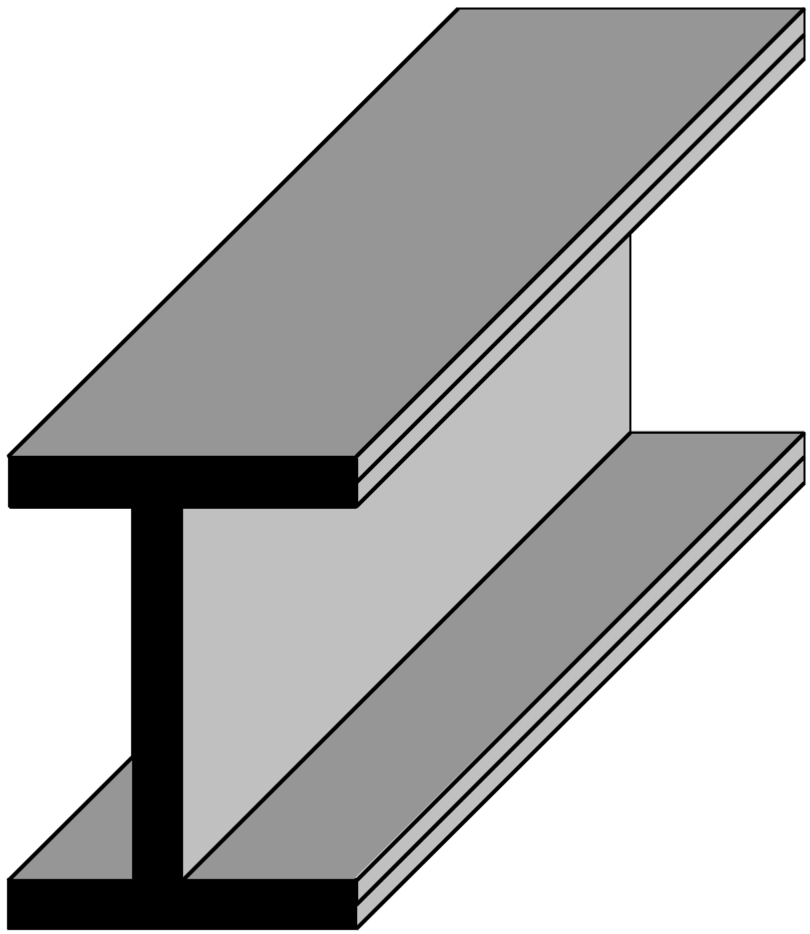
Decision Variables

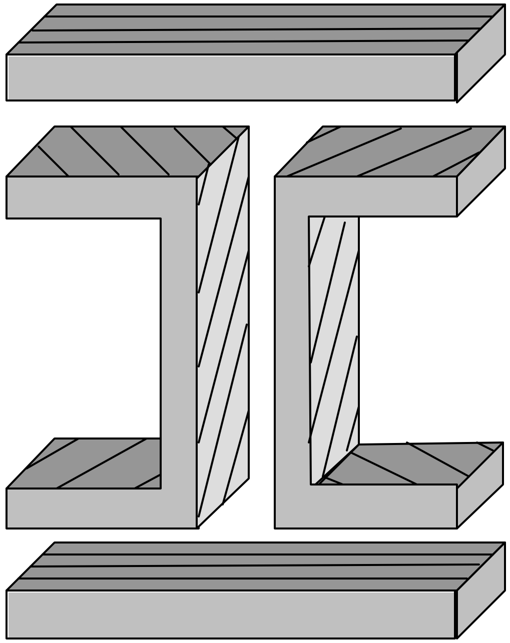


Sandwich Panel



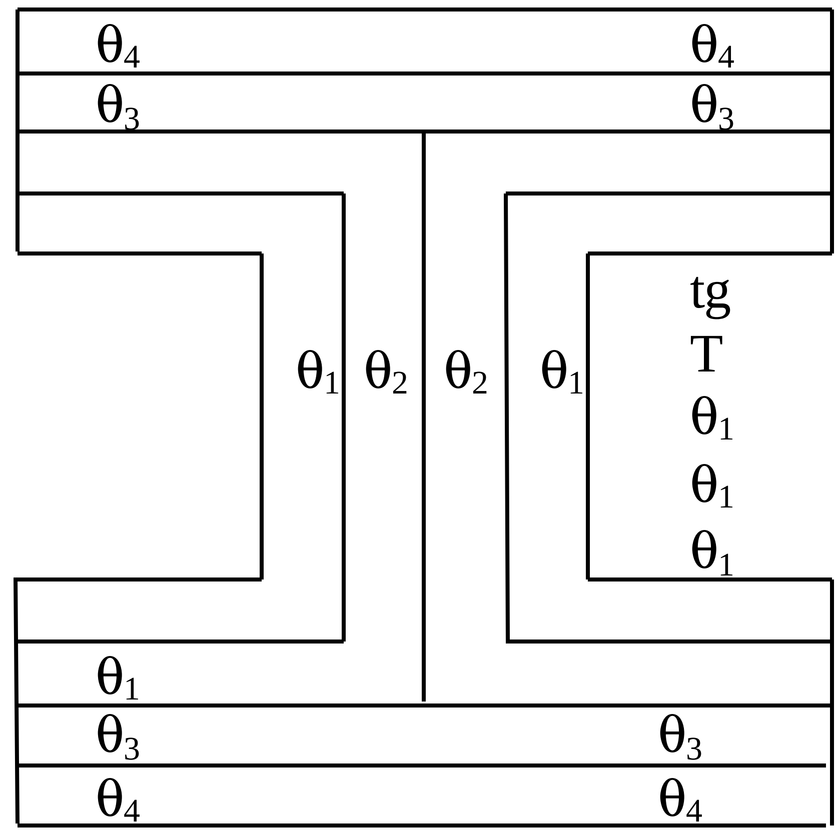
I-Beams





Fiber Angle Decision Variables

Manufacturing Considerations: plies extend through both flanges and the webs, and may cause an abrupt change of fiber angles in flanges



Global Optimization

- Maximize performance, which could be margins of safety associated with strain, stiffness, strength, and buckling analyses
  - maximize $f_{stiffness}$
  - maximize $\min\{f_{\text{margin of safety}}\}$

[Graesser, Zabinsky, Tuttle, Kim, Composite Structures 18, 1991]

Stiffness of Laminate

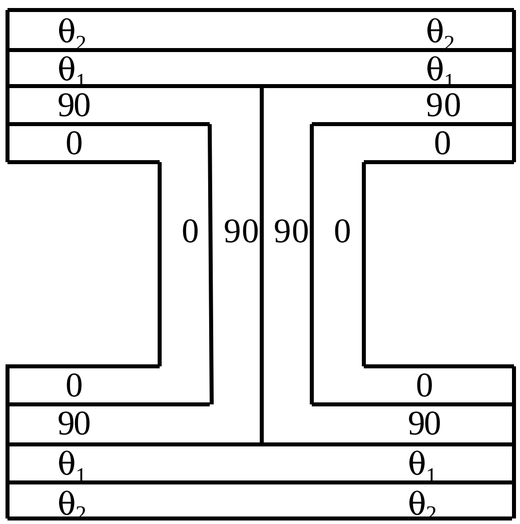
- 4 ply symmetric laminate (θ_1, θ_2, θ_2, θ_1)

$f_{stiffne}$

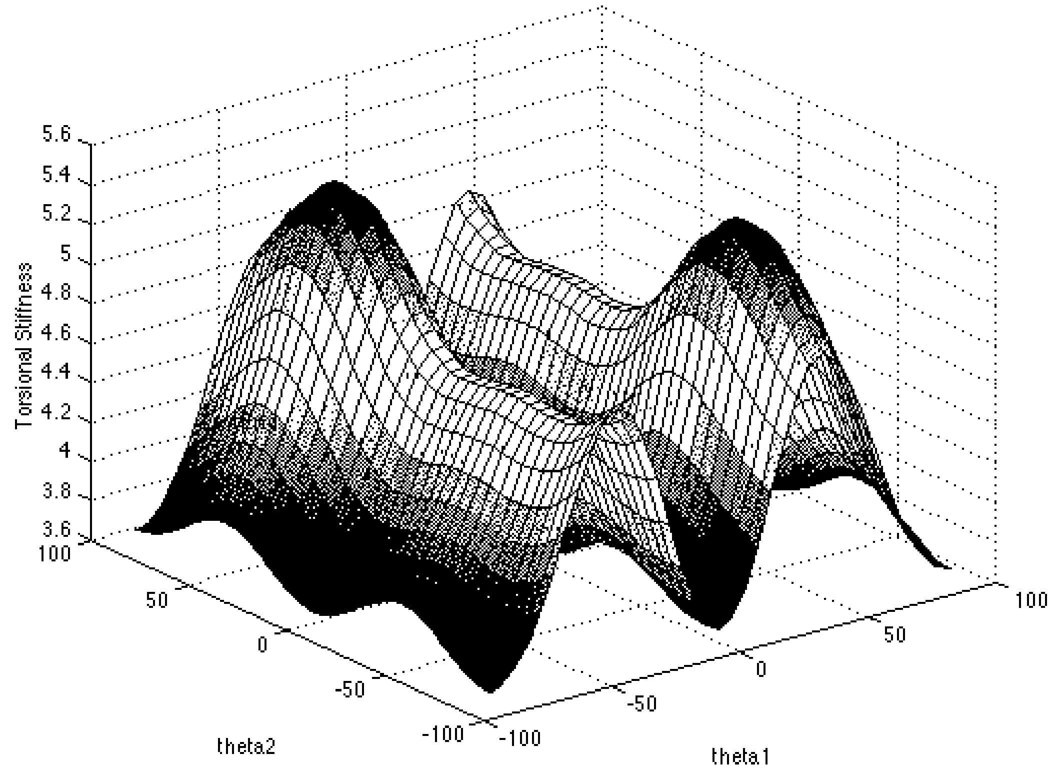
critical$_{sti}$

θ_1
θ_2
θ_2
θ_1

Beam Stiffness Function



Four ply symmetric beam



Optimization Formulation of I-beam

max axial beam stiffness

s.t. torsional beam stiffness $\geq$ equivalent Al - beam torsional stiffness

bending beam stiffnesses $\geq$ equivalent Al - beam bending stiffnesses

longitudinal wall stiffness $\geq$ 2 times transverse ply stiffness

transverse wall stiffness $\geq$ 2 times transverse ply stiffness

$-90^0 \leq$ fiber directions $\leq 90^0$

[Savic, Tuttle, Zabinsky, Composite Structures 53, 2001]

Hierarchical Multi-objective Formulation

- Minimize f_{weight}
  Maximize $\min\{f_{margin\ of\ safety}\}$

- Subject to:

$$f_{margin\ of\ safety} \geq 0$$
$$-90 \leq \theta_i \leq +90$$
θ_i takes on discrete values

How can we solve…?

IDEAL Algorithm:

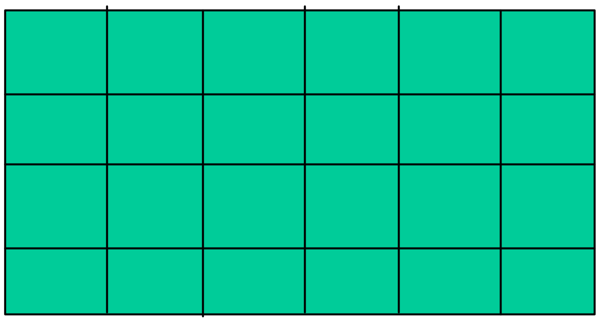
- optimizes any function quickly
- handles continuous and/or discrete variables
- is easy to implement and use

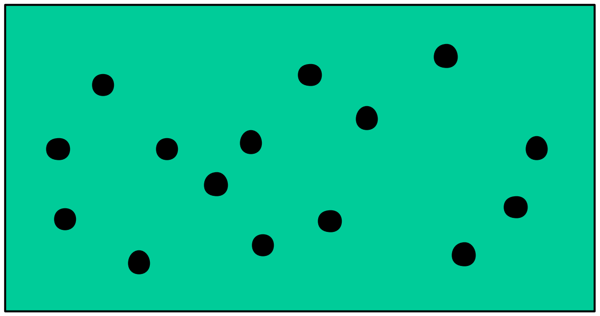
Theoretical Performance of Stochastic Adaptive Search

- What kind of performance can we hope for?
- Global optimization problems are known to be NP-hard
- Sacrifice guarantee of optimality for speed in finding a “good” solution

Two Simple Methods

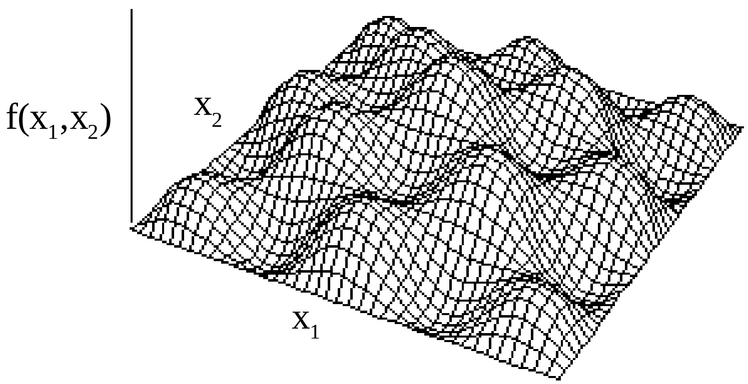
- **Grid Search**: Number of grid points is $O((L/e)^n)$, where L is the Lipschitz constant, n is the dimension, and e is distance to the optimum
- **Pure Random Search**: Expected number of points is $O(1/p(y^*+e))$, where $p(y^*+e)$ is the probability of sampling within e of the optimum y^*
- Complexity of both is exponential in dimension

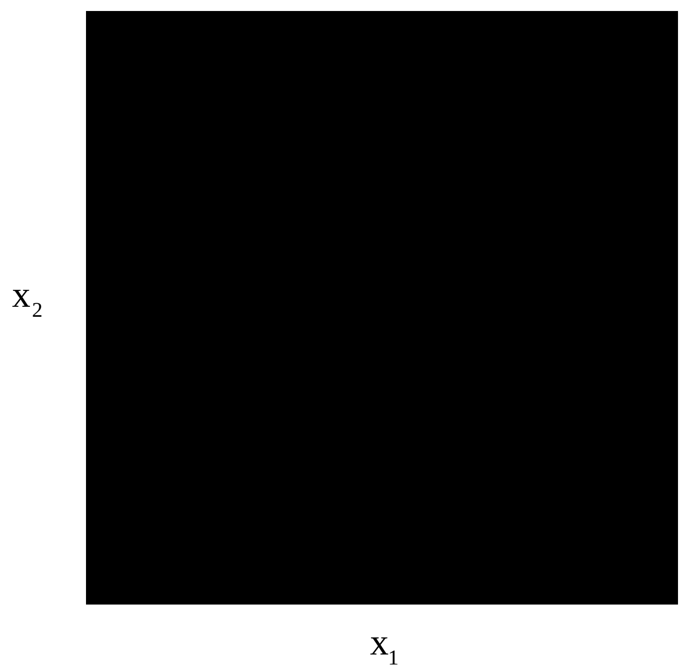




Pure Adaptive Search (PAS)

- PAS: chooses points uniformly distributed in improving level sets





Bounds on PAS

- PAS (continuous):

$$E[N] \leq 1 + ln\,(1/p(y^*+e)\,)$$

where $p(y^*+e)$ is the probability of sampling within e of the global optimum y^*

- PAS (finite discrete):

$$E[N] \leq 1 + ln\,(1/p_1\,)$$

where p_1 is the probability of sampling the global optimum

PAS

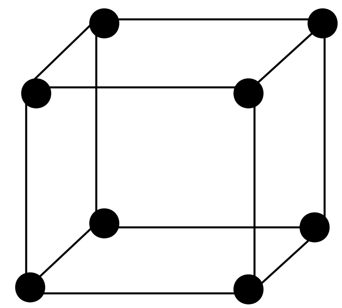
- Theoretically, PAS is LINEAR in dimension
- Theorem:

  For any global optimization problem in n dimensions, with Lipschitz constant at most L, and convex feasible region with diameter at most D, the expected number of PAS points to get within e of the global optimum is:

$$E[N(y^*+e)] \leq 1 + n \ \ln(LD / e)$$

[Zabinsky and Smith, 1992]

Finite PAS



- Analogous LINEARITY result
- Theorem:

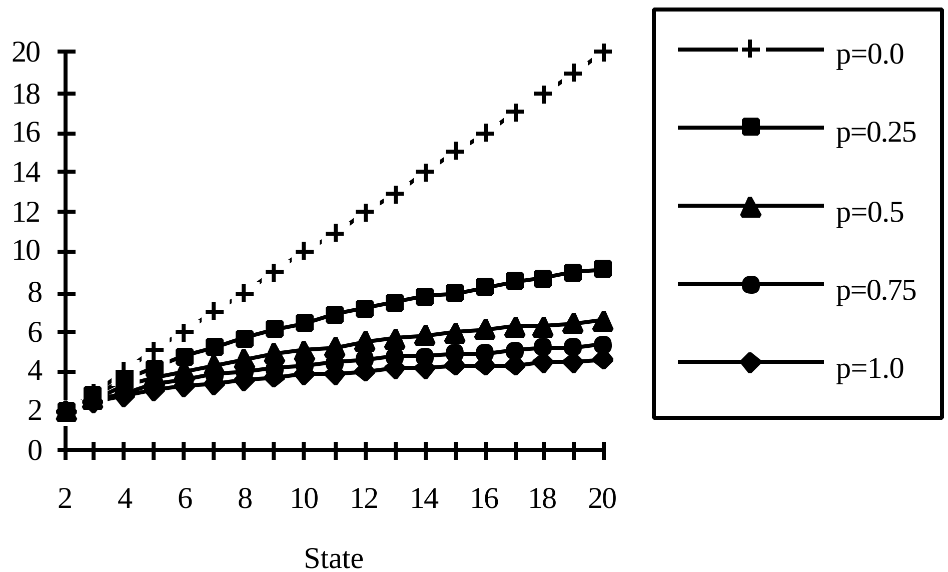
For an n-dimensional lattice $\{1, \ldots, k\}^n$ with distinct objective function values, the expected number of points, sampling uniformly, to first reach the global optimum is:

$$E[N(y^*)] \leq 2 + n \; ln\,(k)$$

[Zabinsky, Wood, Steel and Baritompa, 1995]

Power of Improvement: Combine PAS and PRS

- PAS is difficult to implement directly
- How much worse is the performance if we use PAS with probability *p*, and PRS with probability *1-p* ?



[Zabinsky and Kristinsdottir, 1997]

Hesitant Adaptive Search (HAS)

- What if we sample improving level sets with bettering probability *b(y)* , and "hesitate" with probability *1-b(y)* ?

$$E[N(y^*+e)] = \int_{y^*+e}^{\infty} \frac{dr(t)}{b(t)\,p(t)}$$

where *r(t)* is the underlying sampling distribution and *p(t)* is the probability of sampling *t* or better

[Bulger and Wood, 1998]

General HAS

- For a mixed discrete and continuous global optimization problem, the expected value of $N(y^*+e)$, the variance, and the complete distribution can be expressed using the sampling distribution $r(t)$ and bettering probabilities $b(y)$

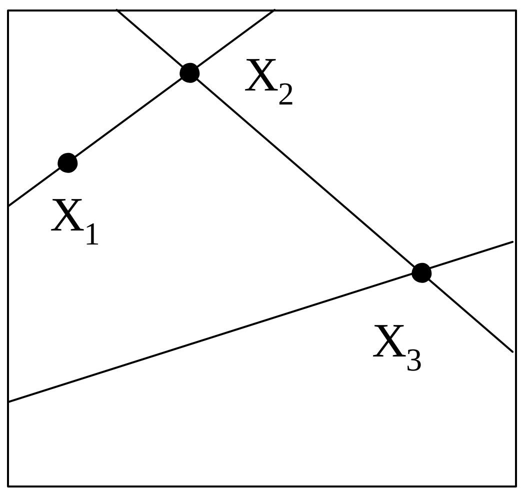
[Wood, Zabinsky, and Kristinsdottir, 2001]

Backtracking Adaptive Search (BAS)

- What if we sometimes accept "worse" points, in order to move across a barrier, and reach the global optimum?
- **Discrete BAS** - use Markov chain analysis, $(I-Q)^{-1}e$, to obtain expected number of iterations [Kristinsdottir, Zabinsky and Wood, to appear]
- **Continuous BAS** - define "worsening probability" and obtain an integral equation [Bulger et al., submitted]

How can we implement PAS?

- To obtain the linearity result, can we consistently generate uniformly distributed points in improving level sets?
- Hit-and-Run can generate asymptotically uniform points [Smith, 1984]



Improving Hit-and-Run (IHR)

- Use Hit-and-Run to generate approximately uniform points within improving level sets
- How long should each Hit-and-Run sequence be?

  if very long (infinite) sequences, then points are approximately uniform, thus a linear number of long (infinite) sequences

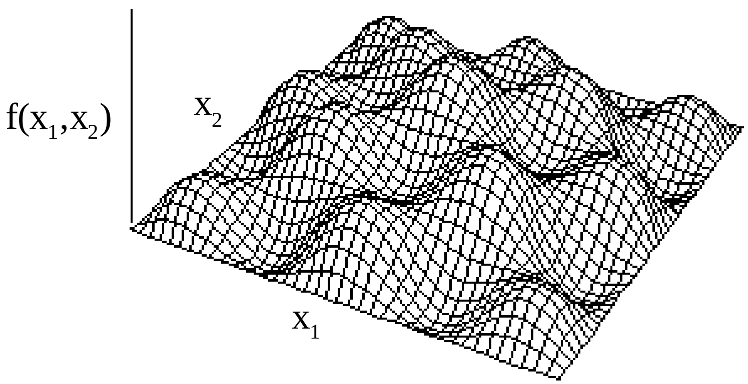
$$(n)(\infty)$$

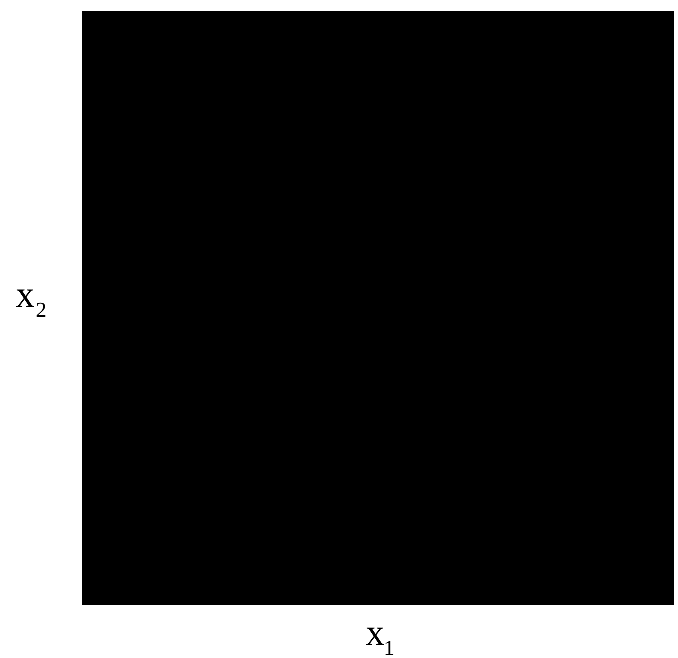
  if very short (1) sequences, then is it efficient?

$$(?)(1)$$

IHR

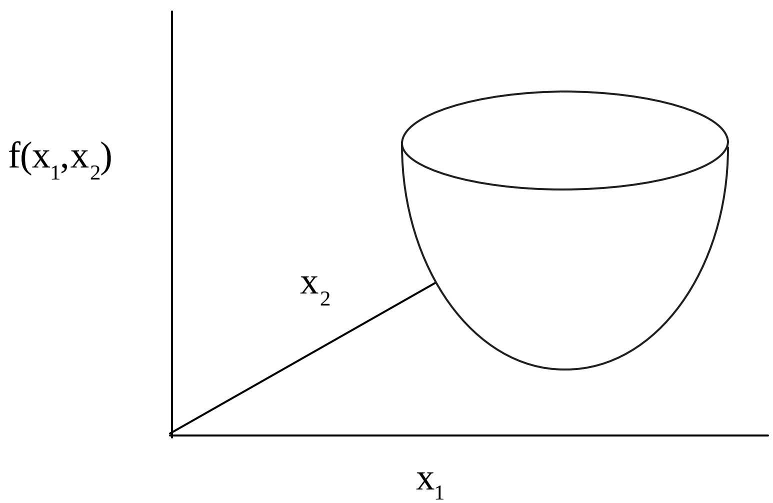
- IHR: choose a random direction and a random point





Is IHR efficient in dimension?

- Theorem:
  For any elliptical program in *n* dimensions, the expected number of function evaluations for IHR is: $O(n^{5/2})$ [Zabinsky, Smith, McDonald, Romeijn, Kaufman, 1993]

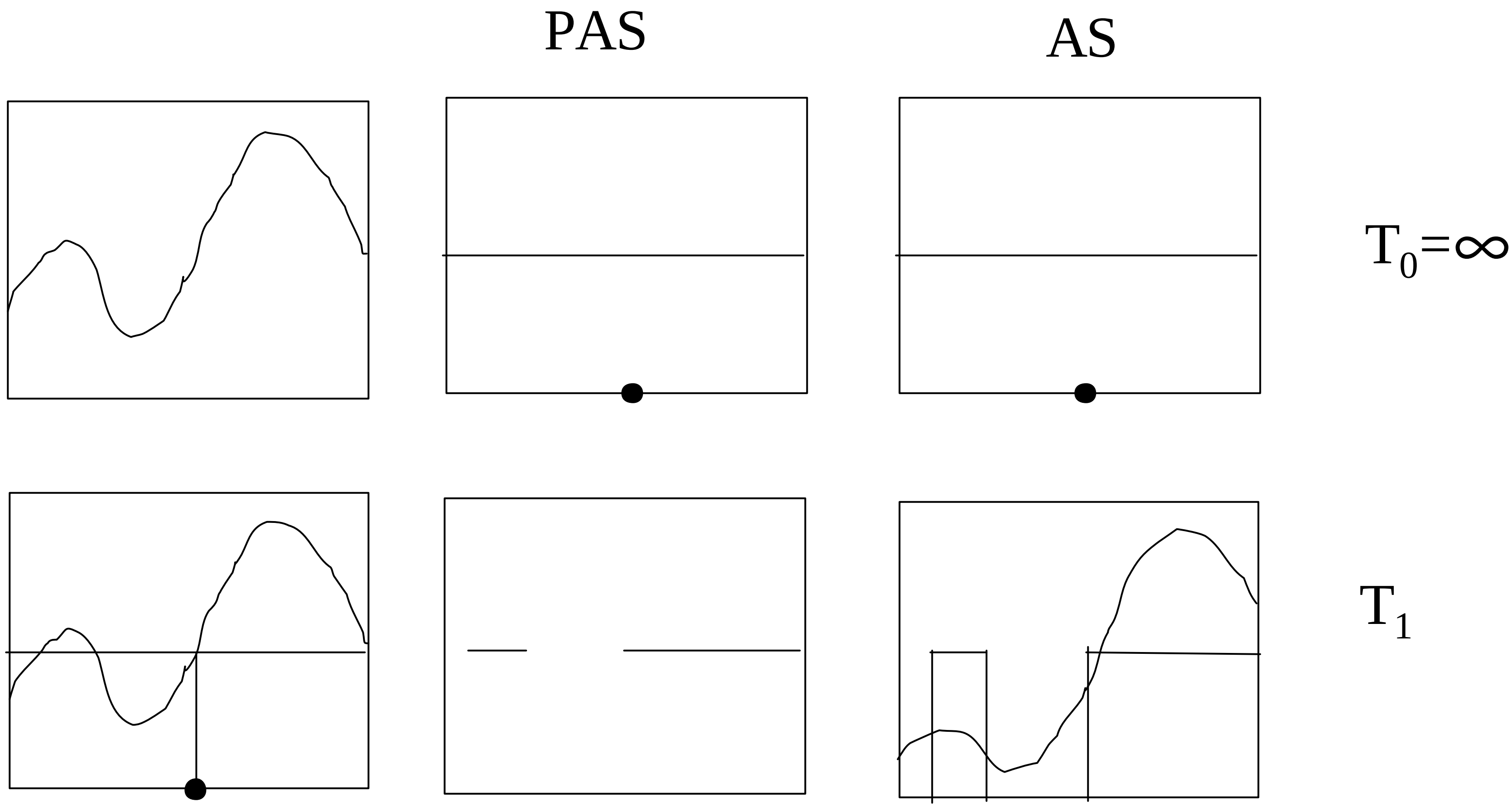


Adaptive Search

- Can we relax PAS slightly and still retain linearity? Yes!
- Adaptive Search:
  - generate points over the whole domain using a Boltzmann distribution parameterized by temperature T, to achieve a probability of 1-a of hitting the improving region

    [Romeijn and Smith, 1994]

Sampling Distributions for PAS and Adaptive Search



Implement Adaptive Search

- Generate points in entire domain according to Boltzmann distribution, instead of in the improving level set
- Modify Hit-and-Run by adding a probabilistic Metropolis acceptance/rejection criterion to approximate the Boltzmann distribution

Hide-and-Seek

- Add an acceptance probability with a temperature and a cooling schedule to the Hit-and-Run generator
  - Given current point X_k, accept candidate point W_k with probability $min\{1, exp(f(X_k)-f(W_k))/T\}$
  - Hit-and-Run with this Metropolis criterion as an acceptance probability with constant temperature converges to a Boltmann distribution with parameter T

[Romeijn and Smith, 1994]

Cooling Schedules for Hide-and-Seek

- Hide-and-Seek will converge in probability, with almost any cooling schedule that drives temperature to zero
- Adaptive cooling schedule, based on a quadratic function and f*;

$$T_{k+1} = 2(f^*-Y_k)/ c^2_{1-a}(n)$$

[Romeijn and Smith, 1994]

Cooling Schedules for Hide-and-Seek

- Use a consistent estimator f' of f*;
  $f'(X_0, \ldots, X_k) = Y_{(k)} - (Y_{(k-1)} - Y_{(k)})/((1-q)^{-n/2} - 1)$
  and adjust at every record value
- Geometric; $T_{k+1} = constant * T_k$
  and adjust every N_T iterations
- IHR; $T_{k+1} = 0$

Modifications to Direction Generator

- Modify the direction choice sampling distribution within Hit-and-Run
  - HD: hyperspherical directions
  - CD: coordinate directions [Berbee, et al., 1987]
  - Reflection Generator: uses hyperspherical directions that bounce off boundary [Romeijn, Zabinsky, Graesser, Neogi, 1999]
  - Artifical Centering Hit-and-Run: non-uniform direction that uses the Hessian information and optimizes the rate of convergence [Kaufman and Smith, 1998]

Compare HD and CD Direction Generators

- Theoretical Comparison:
  - Use a Markov chain analysis to find the expected number of iterations
- Numerical Comparison:
  - Computational results

[Kristinsdottir, 1997]

Markov Chain Analysis

- Compare expected number of iterations until first reaching optimum,

$$E[N(y^*)]=(I-Q)^{-1}e$$

- States are discrete points in domain

	x_1	x_2		$\cdots$		x_*
x_1						$p_{1,*}$
x_2						$p_{2,*}$
$\vdots$						$\vdots$
x_*	0	0	0	$\cdots$	0	1

Test on Discretized Quadratic Problem

- Minimize $f(x_1, x_2) = x_1^2 + x_2^2$ subject to x_1, x_2 in $\{-1,0,1\}$

	-1	2	1	2
x_2	0	1	0	1
	1	2	1	2
		-1	0	1
			x_1	

IHR - CD Generator

- Starting at (1,1), the expected number of iterations until first sampling optimum is

$$E[N(y^*)] = 9.0$$

	−1,1	0,1	1,1	−1,0	1,0	−1,−1	0,−1	1,−1	0,0
−1,1	2/6	1/6	1/6	1/6	0	1/6	0	0	0
0,1	0	4/6	0	0	0	0	1/6	0	1/6
1,1	1/6	1/6	2/6	0	1/6	0	0	1/6	0
−1,0	0	0	0	4/6	1/6	0	0	0	1/6
1,0	0	0	0	1/6	4/6	0	0	0	1/6
−1,−1	1/6	0	0	1/6	0	2/6	1/6	1/6	0
0,−1	0	1/6	0	0	0	0	4/6	0	1/6
1,−1	0	0	1/6	0	1/6	1/6	1/6	2/6	0
0,0	0	0	0	0	0	0	0	0	1

IHR - HD Generator

- Starting at (1,1), the expected number of iterations until first sampling optimum is

$$E[N(y^*)] = 10.4$$

	−1,1	0,1	1,1	−1,0	1,0	−1,−1	0,−1	1,−1	0,0
−1,1	0.446	0.111	0.055	0.111	0.049	0.055	0.049	0.039	0.079
0,1	0	0.703	0	0.072	0.072	0	0.050	0	0.102
1,1	0.055	0.111	0.446	0.049	0.111	0.039	0.049	0.055	0.079
−1,0	0	0.072	0	0.703	0.050	0	0.072	0	0.102
1,0	0	0.072	0	0.050	0.703	0	0.072	0	0.102
−1,−1	0.055	0.049	0.039	0.111	0.049	0.446	0.111	0.055	0.079
0,−1	0	0.050	0	0.072	0.072	0	0.703	0	0.102
1,−1	0.039	0.049	0.055	0.049	0.111	0.055	0.111	0.446	0.079
0,0	0	0	0	0	0	0	0	0	1

Compare HD and CD Analytically and Empirically

- Discretized quadratic problem (20 runs each)

			CD		HD	
Domain	Dim	# states	Model	Empirical	Model	Empirical
3×3	2	3^2	9	12	10	9
	3	3^3	17	20	32	26
	10	3^{10}	–	84	–	273
	30	3^{30}	–	334	–	1657
	50	3^{50}	–	763	–	4330
5×5	2	5^2	15	19	19	16
	3	5^3	27	37	55	36
	10	5^{10}	–	134	–	488
	30	5^{30}	–	616	–	3869
	50	5^{50}	–	1125	–	8824
11×11	2	11^2	33	34	47	32
	3	11^3	61	71	77	104
	10	11^{10}	–	304	–	1343
	30	11^{30}	–	1439	–	8453
	50	11^{50}	–	2512	–	21651
51×51	2	51^2	153	129	228	202
	3	51^3	–	315	–	471
	10	51^{10}	–	1370	–	5526
	30	51^{30}	–	6565	–	(11) 32487
	50	51^{50}	–	11799	–	**
101×101	2	101^2	–	281	–	438
	3	101^3	–	619	–	944
	10	101^{10}	–	2961	–	13422
	30	101^{30}	–	10844	–	(3) 44213
	50	101^{50}	–	24473	–	**

Compare HD and CD on a Global Problem

- Minimize $f(x_1, x_2)$
  subject to x_1, x_2 in $\{-1,0,1\}$

x_2 \ x_1	-1	0	1
-1	1	2	1
0	2	0	2
1	1	2	1

CD Generator Doesn’t Converge!

HD Generator: 12.7 iterations

Try CD Generator with Acceptance Probability

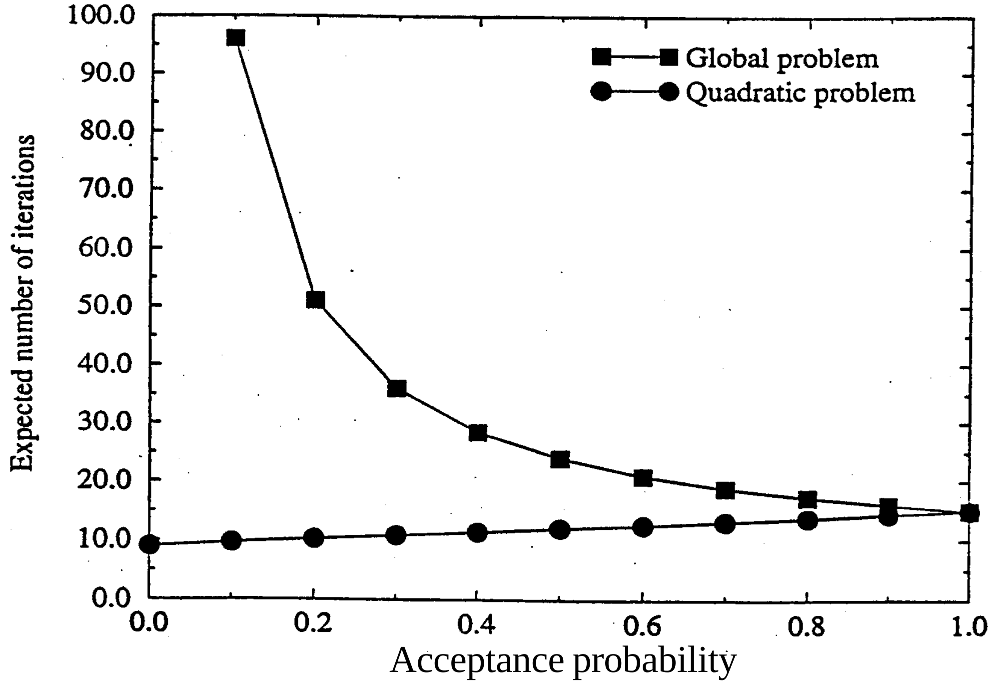
- Add a constant probability *p* of accepting a non-improving point

	$-1,1$	$0,1$	$1,1$	$-1,0$	$1,0$	$-1,-1$	$0,-1$	$1,-1$	$0,0$
$-1,1$									
$0,1$									
$1,1$									
$-1,0$	0	0	0	$4/6$	$1/6$	0	0	0	$1/6$
$-1,0$	$p/6$	0	0	$2/6+2(1-p)/6$	$1/6$	$p/6$	0	0	$1/6$
$1,0$									
$-1,-1$									
$0,-1$									
$1,-1$									
$0,0$									

CD Generator Varying Acceptance Probability

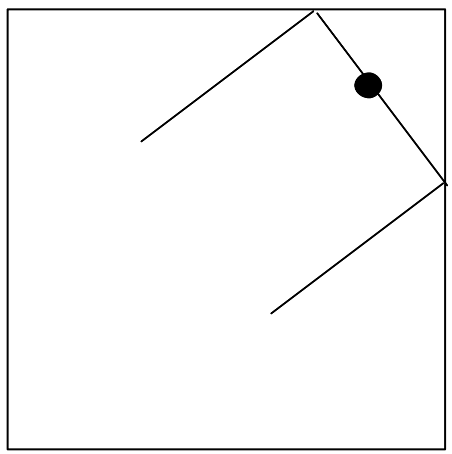
- Expected number of iterations:

	CD	HD
Quadratic problem	9.0	10.4
Global problem	+infinity	12.7



Reflection Generator

- Modification of HD to reduce jamming when in a corner or close to a boundary
- If reflection generator gives a well-defined transition density, the algorithm will converge w.p.1



[Romeijn, et al., 1999]

Optimal Direction Choice

- Choose a non-uniform direction distribution that optimizes the rate of convergence of Hit-and-Run to its target distribution
- Best choice is based on knowing the Hessian at the optimum
- Artificial Centering Hit-and-Run is a heuristic adaptive direction choice rule based on this optimal choice
  [Kaufman and Smith, 1998]

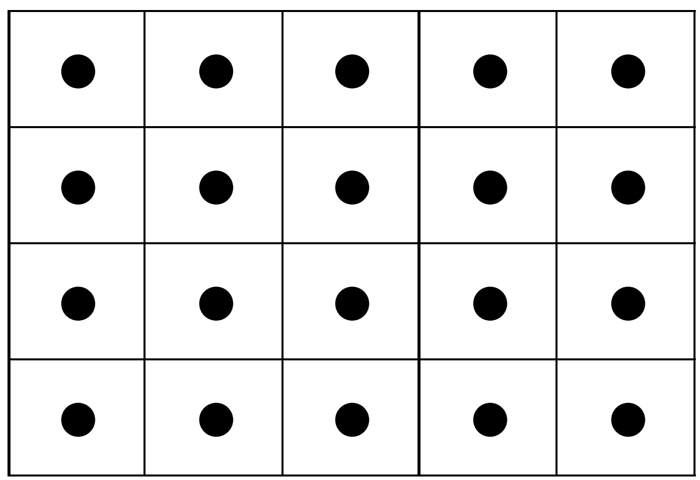
Adapt Hit-and-Run to Discrete

- Modify Hit-and-Run generator to work in a mixed domain of continuous / discrete variables
- Use a step function approach

[Romeijn, Zabinsky, Graesser, Neogi, 1999]

Discretized Version

- When generating the next point, either round to the nearest discrete value, or use its value creating a step function
- Convergence with probability 1 to a broad class of problems



[Romeijn, Zabinsky, Graesser and Neogi, 1999]

Does it work on engineering design problems?

Applications to Aircraft Design

- Applied to the design of a crown panel
  [Swanson, et al., 1991]
- Applied to the design of a keel panel
  [Mabson, et al., 1994]
- Applied to the design of a window belt
  [Metschan, et al., 1994]
- Applied to the design of a full barrel
  [Neogi, 1997]
- Applied to the design of I-beams
  [Savic, et al., 2001]

Manufacturing Tolerances

- Manufacturing processes are not able to reproduce the optimal design exactly, and may have $\pm\delta$ tolerances
- The intended design may be feasible, but the actual design within $\pm\delta$ may fail
- We seek a near-optimal design with feasible $\pm\delta$ tolerances

Illustrate Tolerance Box

- Optimization problem

minimize $f(x) = \text{weight}$

s.t.

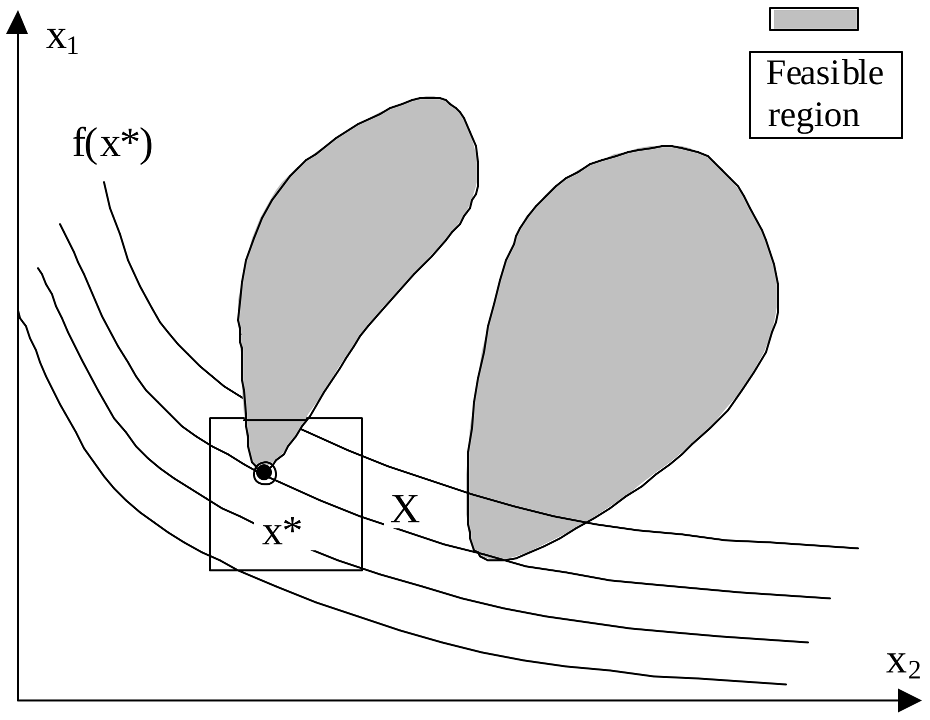
$g_j(x) = \text{margin of safety} \geq 0$

for $j=1,\ldots,m$

Optimum is x^*, $f(x^*)$

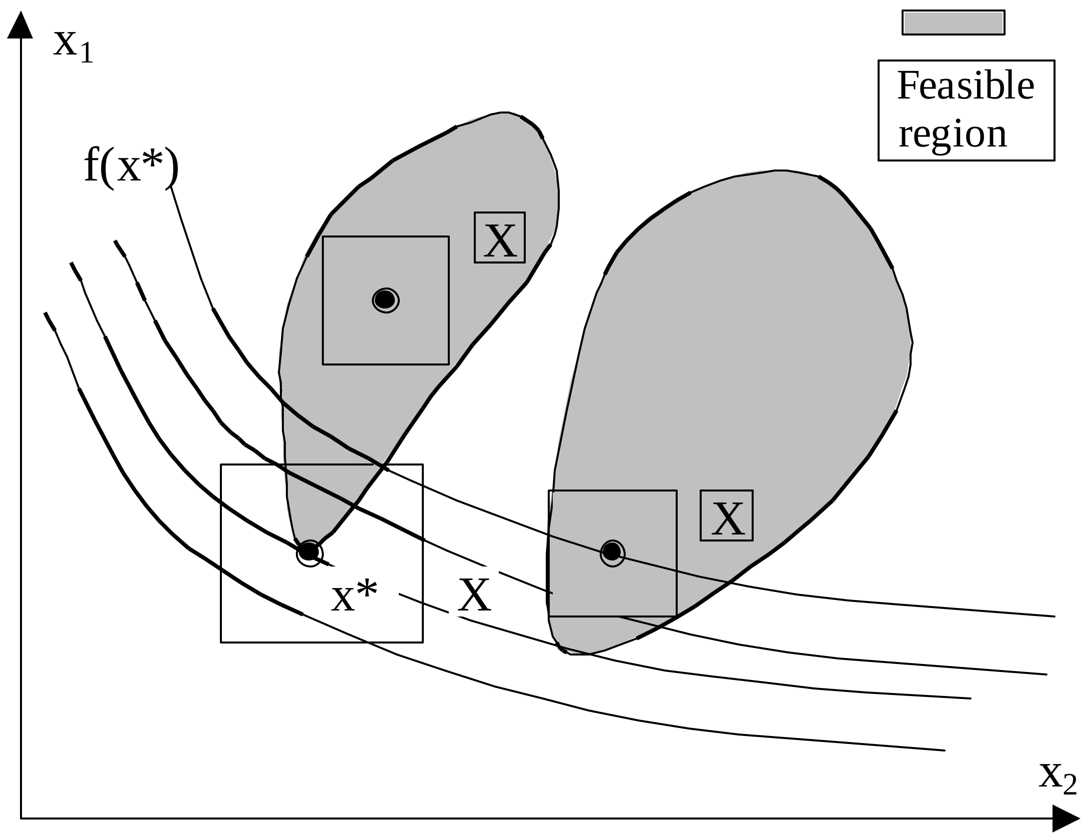
Tolerance box is

$X = [x^*_i-\delta, x^*_i+\delta]$



Find a Near-Optimal Feasible Tolerance Box

- Find a design with $\pm\delta$ tolerances
- Tradeoffs between weight, margin of safety, and size of tolerances

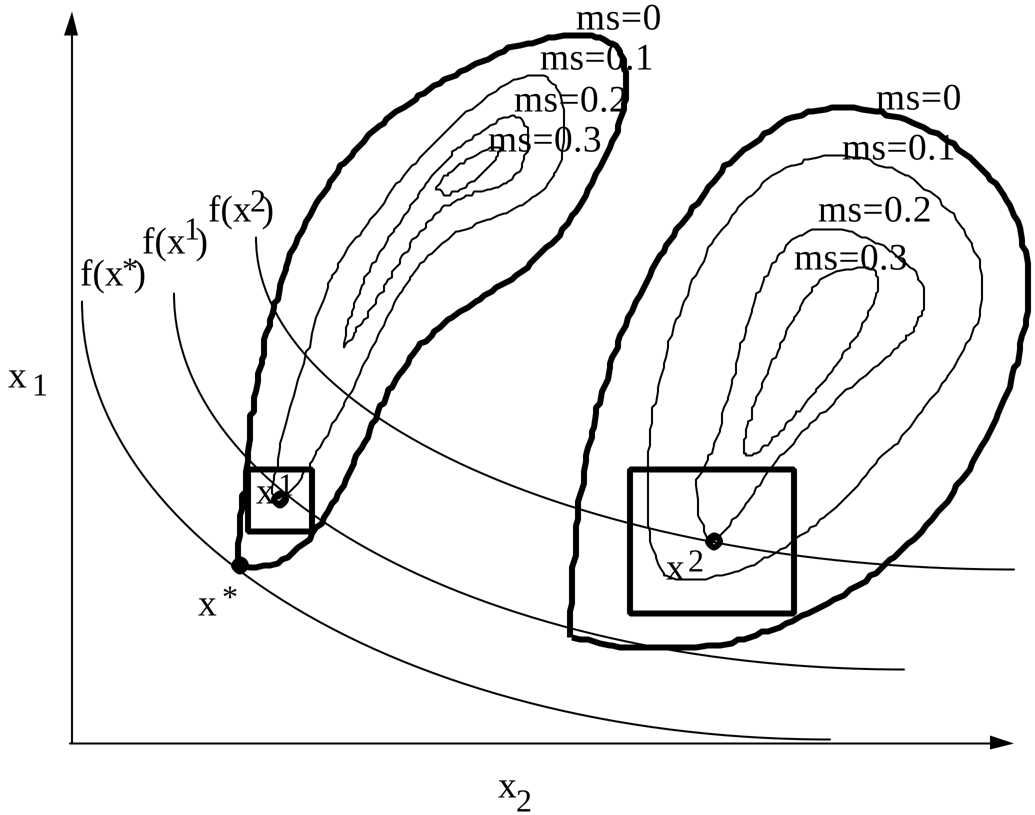


Methodology

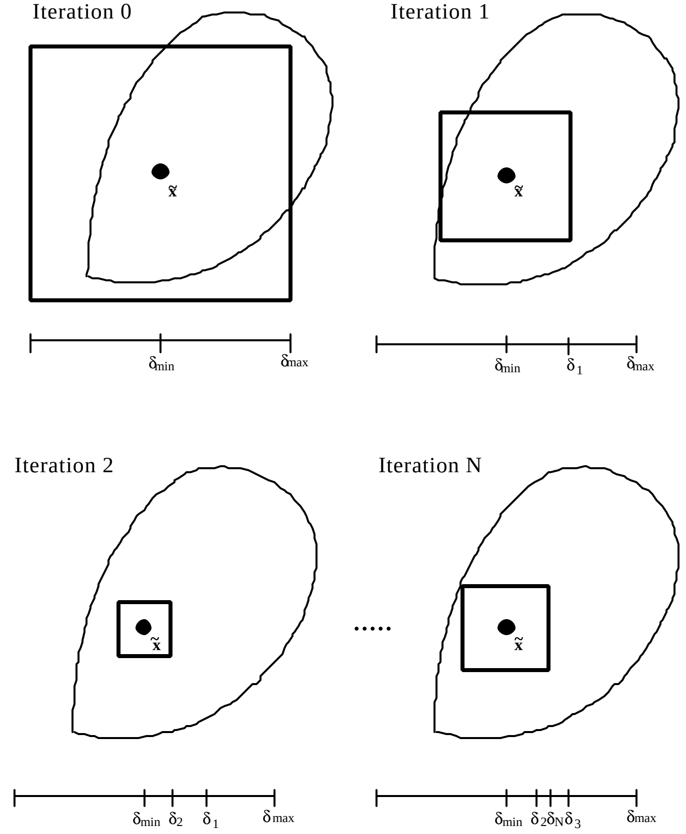
- Use Improving Hit and Run algorithm to optimize (minimize weight)
- Use interval optimization to analyze feasibility of a tolerance box
- Iteratively:
  - Raise the margin of safety
  - Find an optimal solution
  - Analyze feasibility of tolerance box

[Kristinsdottir, Zabinsky, Tuttle, Csendes, 1996]

Multi-objective: weight, tolerances, and margin of safety



Analyze Tolerance Box

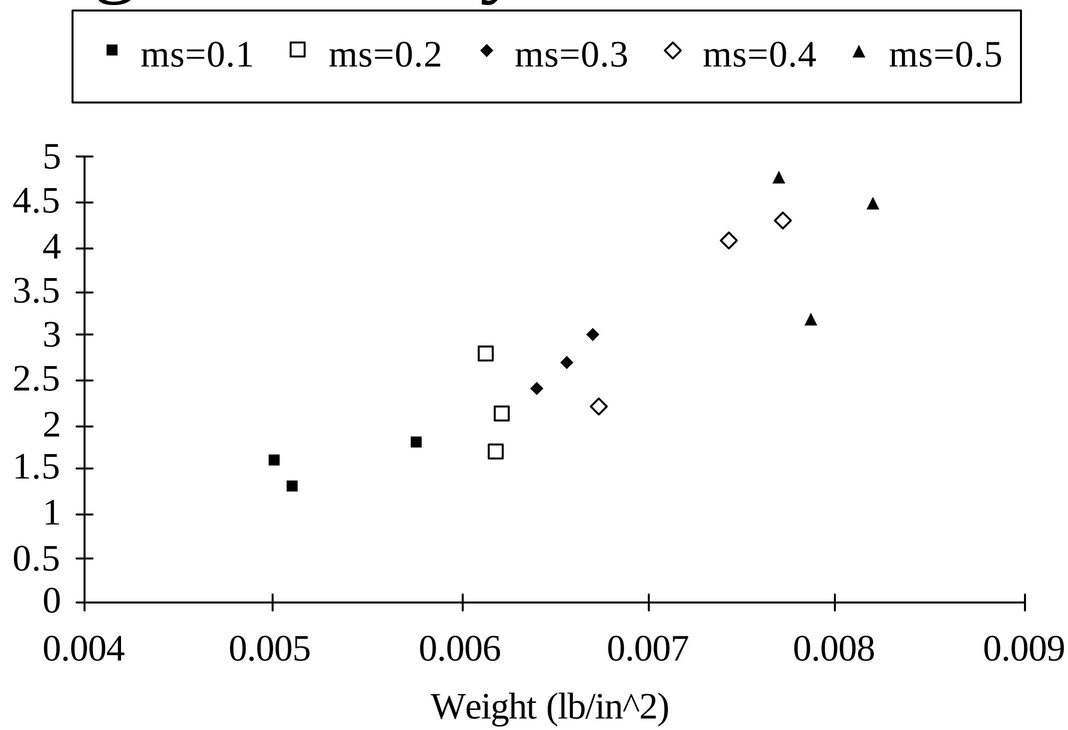


Analyze Tolerances: Check Box

- Interval checking routine:
  - Can guarantee feasibility, may be overcautious
- IHR checking routine:
  - Can guarantee infeasibility, may miss an infeasible point
- Hybrid algorithm:
  - Combines IHR to quickly identify infeasibility, with interval optimization which uses partitioning to guarantee feasibility

Tradeoff Study

- Identify dominating designs
- Consider tradeoffs between weight, tolerances, and margin of safety



Summarize Manufacturing Tolerances

- Developed optimization methodology using interval methods and IHR to consider manufacturing tolerances during preliminary structural design
- Engineer can evaluate tradeoffs between weight, tolerances, and performance
- Engineer can select the appropriate design from a set of dominating designs