THE THEORY AND APPLICATIONS OF DISCRETE LAGRANGE-MULTIPLIER OPTIMIZATION

Benjamin W. Wah

Department of Electrical and Computer Engineering
and the Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801, USA
b-wah@uiuc.edu
http://manip.crhc.uiuc.edu

Contributors: Prof. Y. N. Chang, Prof. Y. Shang, T. Wang, and Z. Wu

Motivations

- Abundant applications in nonlinear discrete constrained optimization
  - Mathematics
  - Machine vision
  - Robotics
  - Database systems
  - Text processing
  - Computer graphics
  - Security
  - Computer science and artificial intelligence
  - Integrated circuit design automation
  - Computer architecture design
  - Computer networks
- NP-hard
  - Complete methods cannot handle large problem instances
  - Heuristic methods have difficulties with nonlinear constraints

Simple Example of Discrete Constrained Optimization

- A nonlinear, non-convex, discrete constrained optimization

- One dimension, one constraint function

$$\min \quad f(x) = sin(2 - 0.4x - 2x^2 + 0.75x^3 + 0.4x^4 - 0.15x^5 + sin(5.0x))8.0 + 0.4x^2$$

subject to

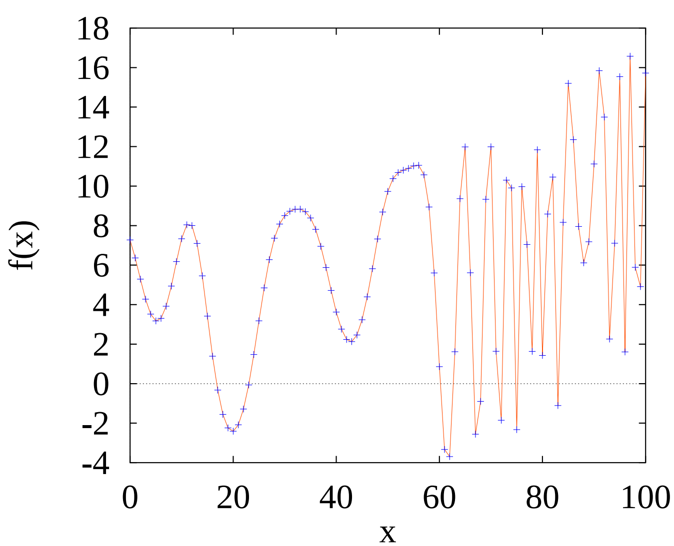
$$g(x) = ||sin(x) - (sin(cos(x)10)log(x))sin(x^2 - 14.0)| - 1.3| - 0.3 \leq 0$$

discrete variable $x : \{0, 1, 2, \cdots, 100\}$

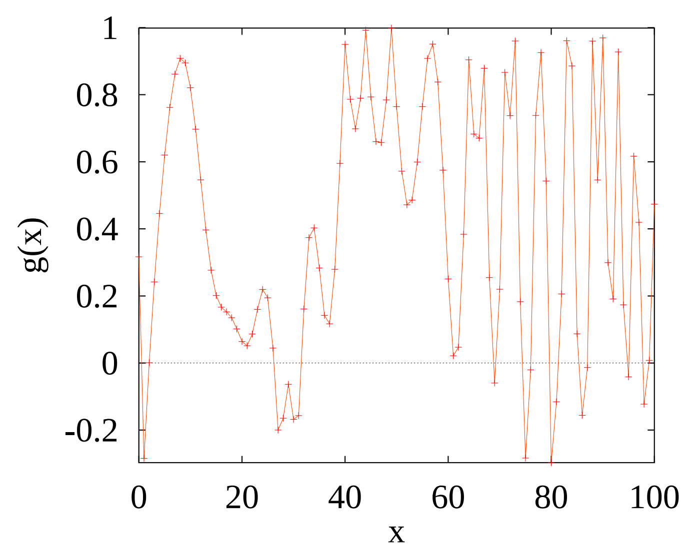
- Optimum solution: $f = -2.325$ at $x = 75$

Simple Example (cont'd)

Objective function



Constraint function



- Feasible points: $\{1, 27, 28, 29, 30, 31, 69, 75, 76, 80, 81, 86, 87, 95, 98\}$

A More Complex Benchmark G2

$$\begin{aligned}
\text{maximize} \quad & \left| \frac{\sum_{i=1}^{n} cos^4(x_i) - 2\prod_{i=1}^{n} cos^2(x_i)}{\sqrt{\sum_{i=1}^{n} i x_i^2}} \right| \\
\text{such that} \quad & \prod_{i=1}^{n} x_i \geq 0.75 \qquad \text{and} \qquad \sum_{i=1}^{n} x_i \leq 7.5n \\
\text{where} \quad & n = 20
\end{aligned}$$

Continuous version

- EA solution: 0.803553
- SQP solution: 0.640329
- CSA solution: 0.803619

Mixed and discretized versions

- CSA solution: 0.803619

Satisfiability (SAT) Problems

- A nonlinear discrete constrained decision problem
  - Given m clauses $C_1, C_2, \cdots, C_m$ in disjunctive form on n variables

    $$x = (x_1, x_2, \cdots, x_n) \qquad x_i \in \{0, 1\}$$

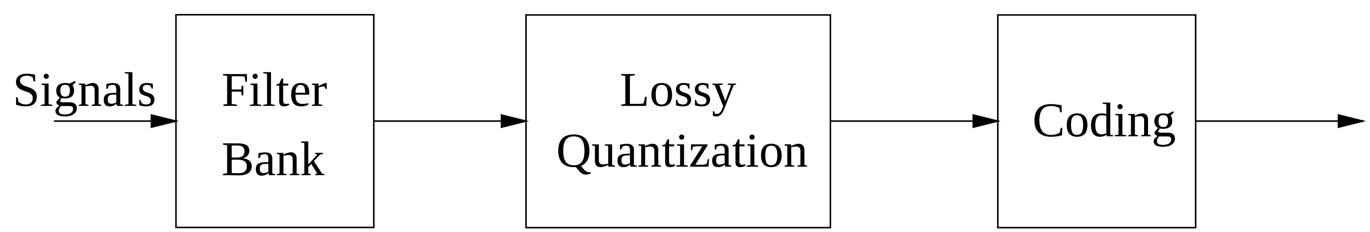
    and a Boolean formula in conjunctive normal form (CNF)

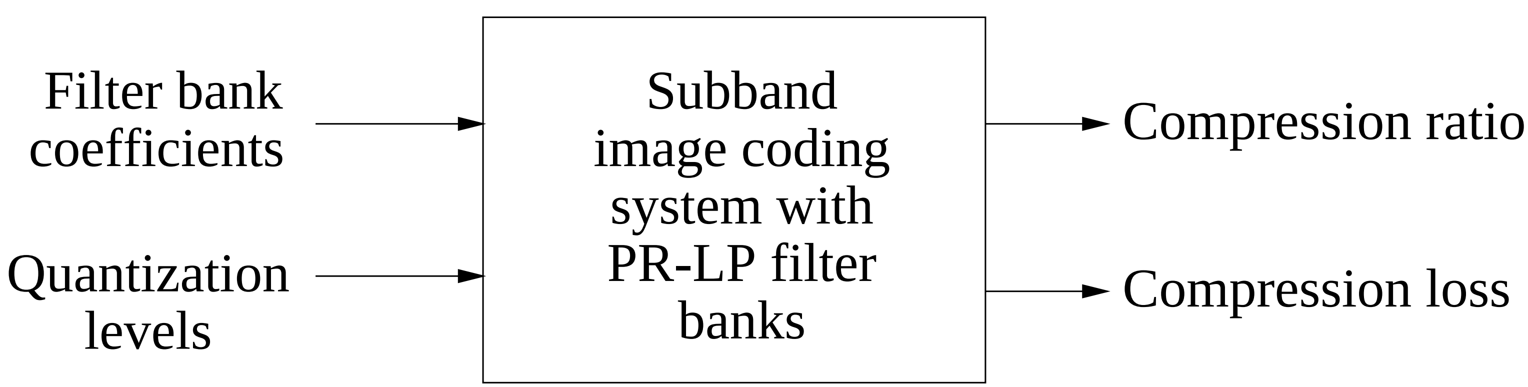
    $$C_1 \wedge C_2 \wedge \cdots \wedge C_m,$$

    find a truth assignment or derive infeasibility.
- Example with 4 variables: $\{x_1,\ x_2,\ x_3,\ x_4\}$
  - 7 clauses: { $\overbrace{(1\ 3\ 4)}^{(x_1 \vee x_3 \vee x_4)}$, $\overbrace{(1\ \text{-}2\ \text{-}3)}^{(x_1 \vee \overline{x_2} \vee \overline{x_3})}$, (-1 -2 4), (-1 -3 -4), (2 3 -4), (2 -3 4), (-2 3 -4) }
  - 2 solutions: { (1 0 0 0), (0 0 1 1) }
- Extension to maximum satisfiability problems

Subband Image Coding

Nonlinear mixed-integer multi-objective constrained optimization problem with black-box model





Nonlinear Discrete Constrained Optimization Problems

- Nonlinear discrete constrained minimization problem

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & g(x) \le 0 & x = (x_1, x_2, \ldots, x_n) \\ & h(x) = 0 & \quad \text{is a vector of discrete variables} \end{array}$$

- Assumptions in discrete Lagrange multiplier theory

1. Feasible local minimum exists
   - Variable space needs not be bounded
   - $f(x)$ should be lower bounded ($\Rightarrow$ existence of minimum)
   - Constraint functions need not be bounded
2. Objective and constraint functions do not need to be differentiable

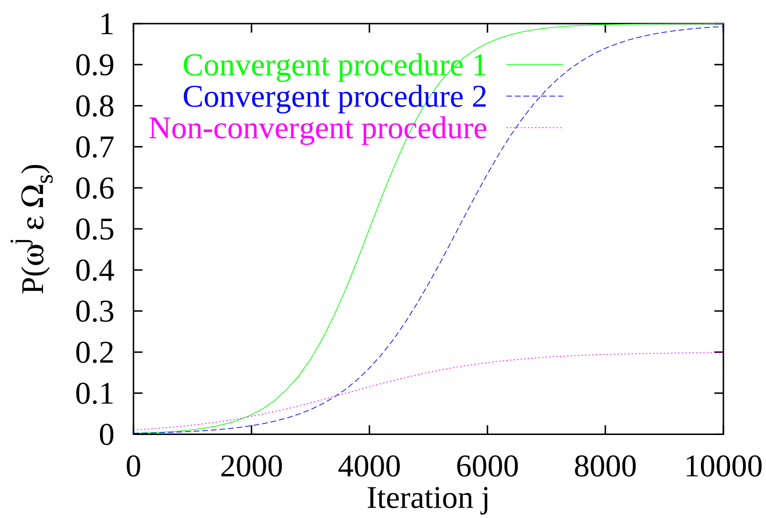
Outline

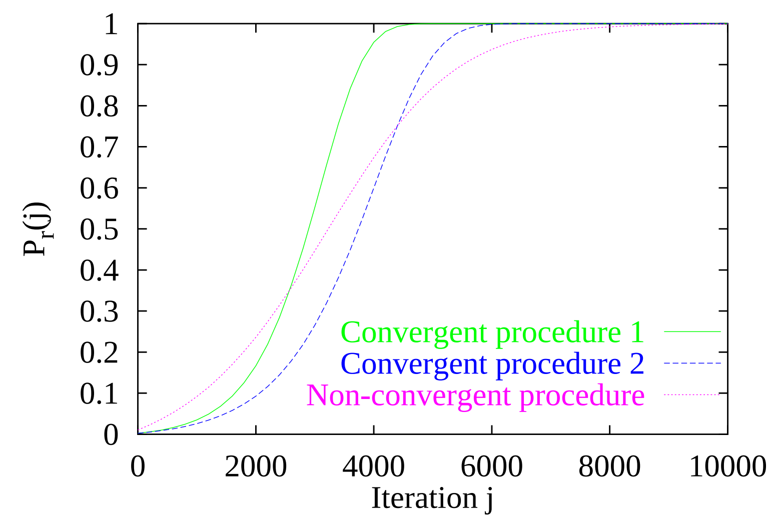
- Previous work on nonlinear discrete constrained optimization
- Theory of discrete Lagrange-multiplier optimization
  - Basic concepts
  - Discrete-space first-order necessary and sufficient conditions
- Extension of theory to continuous and mixed-integer constrained optimization
- Search algorithms
  - DLM: an implementation of the first-order search method
  - CSA: constrained simulated annealing with asymptotic convergence

Previous Work on Nonlinear Discrete Constrained Optimization

- Analytic methods: only for trivial cases
- Decomposition methods are enumerative methods
  - *Branch-and-bound*: works only when constraints can be linearized
  - *Interval method*: works for differentiable continuous functions
  - *Generalized Benders decomposition* and *outer approximation*: works for MINLPs when number of decomposed convex subproblems is manageable
- Stochastic methods: sample variable space by some probability distributions
  - Can at best converge to a constrained global minimum with probability one when time approaches infinity
  - May perform better than enumerative methods under finite time

Asymptotic Convergence in Stochastic Methods





Reachability probability: $P_r(t) = 1 - \prod_{j=1}^{t}(1 - P(\omega^j \in \Omega_s))$

- Asymptotically convergent algorithms always have better reachability probabilities when number of iterations is large
- Many existing stochastic algorithms with reachability
- Open problem on asymptotically convergent algorithm for solving discrete constrained NLP

Previous Work (cont'd)

- Penalty methods: Transformations into unconstrained problems using penalty formulations
  - Static-penalty methods using sufficiently large but finite penalties
    - Suitable penalties are hard to choose
  - Dynamic-penalty methods
    - Sequence of subproblems with increasing static penalties
    - Asymptotic convergence if each subproblem is solved optimally
    - Difficult to solve subproblems optimally, given finite time
  - Lagrange-multiplier methods
    - Requires differentiability and continuity of functions
    - Not for discrete global optimization
  - Lagrangian relaxation: Reformulation of a linear integer minimization problem and solution of resulting problem by Lagrangian duality theory

Previous Work (cont'd)

- Other transformation methods
  - Transformations into constrained 0-1 programming
    - Linearization - only work for simple nonlinear constraints
    - Branch & bound search - fail for high-dimensional nonlinear functions
    - Cutting-plane methods - only work for a few cases
    - Algebraic methods - only work for simple functions
    - Issue: very limited in applicability
  - Transformations into continuous problems
    - Issue 1: continuous solutions may not be feasible in discrete space
    - Issue 2: may not work without differentiability of functions

Outline

- Previous work on nonlinear discrete constrained optimization
- Theory of discrete Lagrange-multiplier optimization
  - Basic concepts
  - Discrete-space first-order necessary and sufficient conditions
- Extension of theory to continuous and mixed-integer constrained optimization
- Search algorithms
  - DLM: an implementation of the first-order search method
  - CSA: constrained simulated annealing with global convergence

Nonlinear Optimization Problems with Equality Constraints

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h(x) = 0 \end{aligned}$$

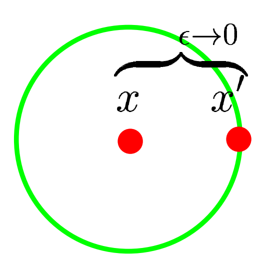
$$x = (x_1, x_2, \ldots, x_n)$$
$$h(x) = (h_1(x), \ldots, h_m(x))$$

Continuous Space	Discrete Space
x is a vector of continuous variables	x is a vector of discrete variables

Neighborhood $N(x)$ of Point x

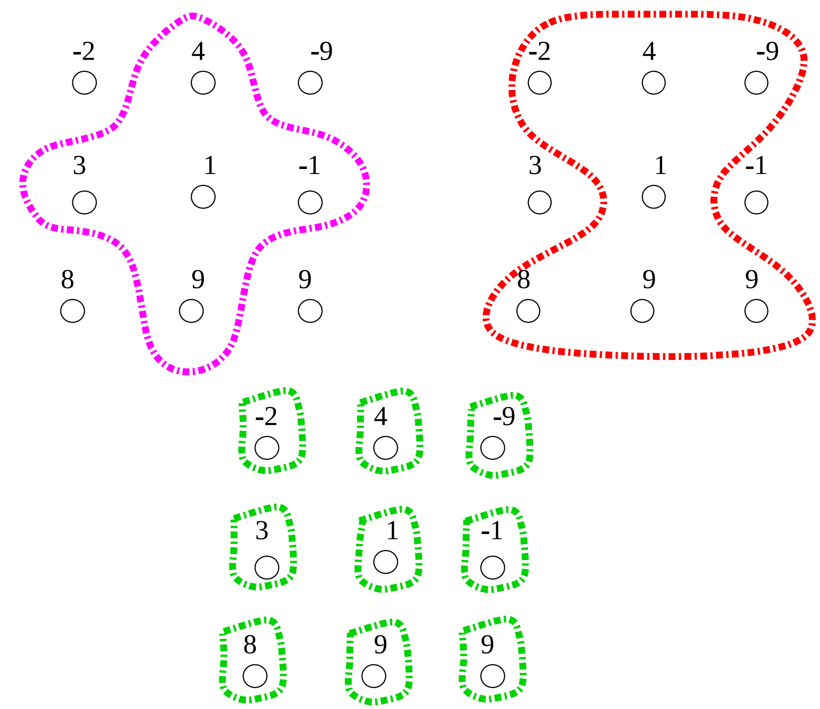
Continuous Space

Defined by open sphere



Discrete Space

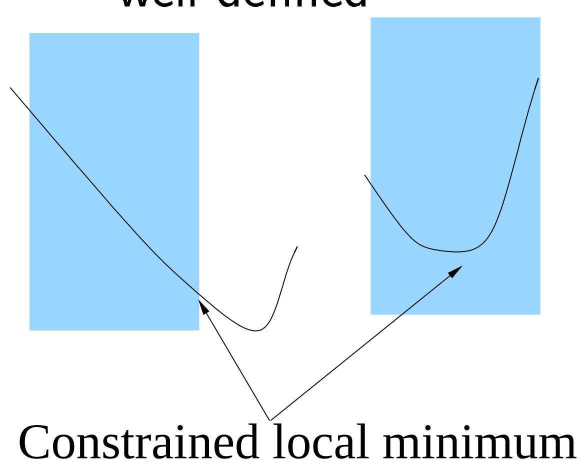
User defined



Constrained Local Minimum (CLM)

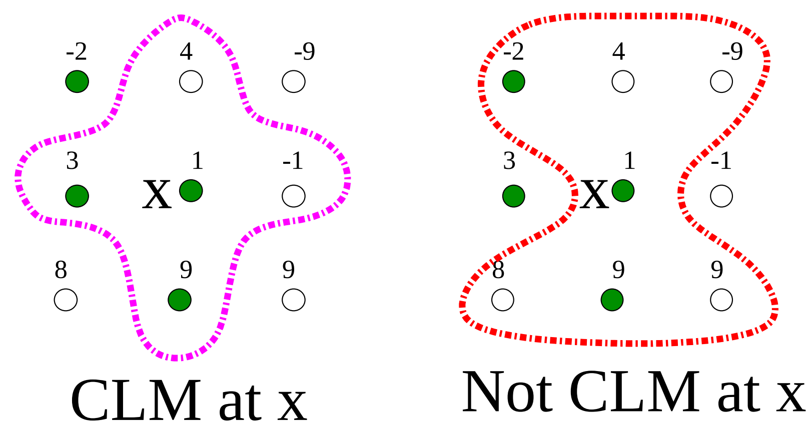
Continuous Space

- Feasible local minimum when compared to feasible points inside an open sphere
- Whether a point is a CLM is well defined



Discrete Space

- Feasible local minimum with respect to neighboring feasible points
- Whether a point is a CLM depends on $N(x)$



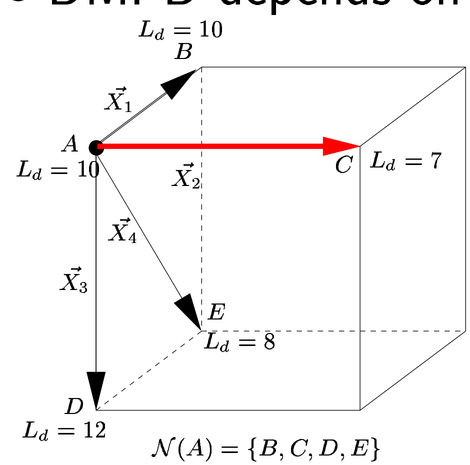
Descent Directions

Continuous Space

- Gradient (∇) provides direction of descents
- Composition, addition, and multiplication of gradients
- Chain rule

Discrete Space

- Direction of Maximum Potential Drop (DMPD)
- $\Delta(x)$ to be a vector that points in the direction of maximum function value drop in the neighborhood of x
- No similar operations on DMPDs
- DMPD depends on $N(x)$



Theory of Lagrangian Multipliers

Continuous Space

- $L_c(x,\lambda) = f(x) + \lambda^T h(x)$
- First-order necessary conditions
  - If x^* is a CLM & *regular point*, then $\exists\ \lambda$ such that

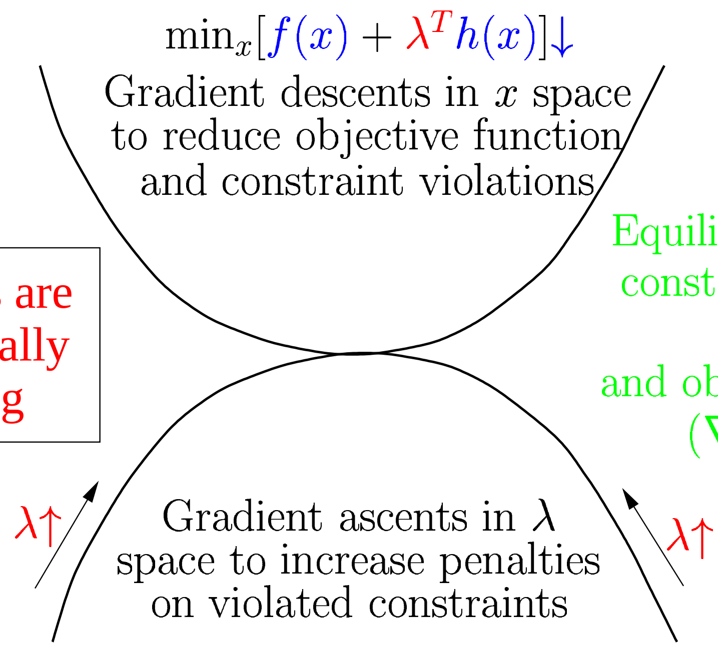
    $$\nabla_x L_c(x,\lambda) = 0, \ \ \nabla_\lambda L_c(x,\lambda) = 0$$
  - Proved using implicit function theorem and chain rule
- Second-order sufficient conditions

Discrete Space

- $L_d(x,\lambda) = f(x) + \lambda^T H(h(x))$
- First-order necessary and sufficient conditions
  - x^* is a CLM, iff
    * x^* satisfies the saddle-point condition or
    * $\exists\ \lambda$ such that

      $$\Delta L_d(x,\lambda) = 0, \ \ h(x) = 0$$
  - Proved through the concept of discrete saddle points

Intuitive Meaning Behind Lagrangian Search



Penalties are dynamically varying

Equilibrium point where constraints are satisfied $(h(x) = 0)$ and objective is minimum $(\nabla_x L(x, \lambda) = 0)$

Saddle Points

Continuous Space

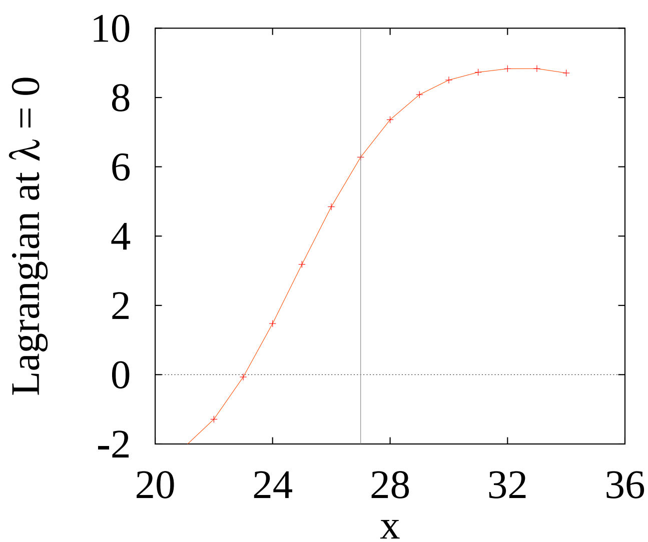
- $L_c(x^*, \lambda) \le L_c(x^*, \lambda^*) \le L_c(x, \lambda^*)$ $\forall \lambda$, x sufficiently close to (x^*, λ^*)
- Not useful concept in continuous space
  - A point can be verified to be a saddle point after it is found
  - No systematic way to find saddle points

Discrete Space

- $L_d(x^*, \lambda) \le L_d(x^*, \lambda^*) \le L_d(x, \lambda^*)$ $\forall (x^*, \lambda)$ and $(x, \lambda^*) \in N(x^*, \lambda^*)$
- Feasible point + local minimum of function $L(x, \lambda^*)$ in x space
- Core of discrete Lagrange-multiplier theory

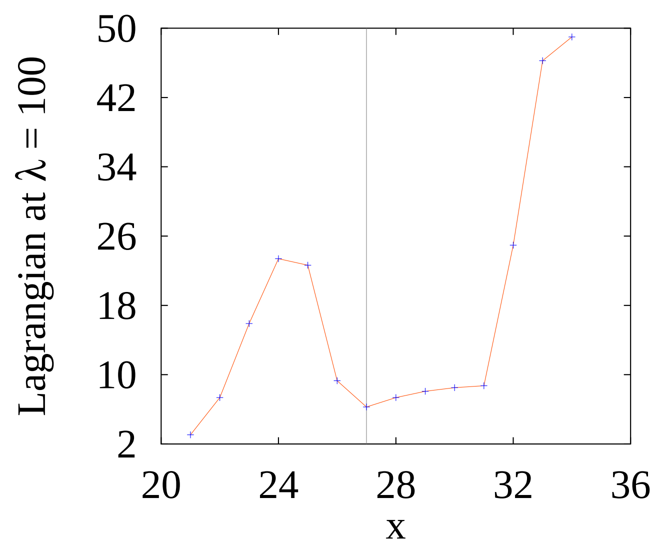
Illustration of Discrete Saddle Points

- $x = \{27, 28, 29, 30, 31\}$ is a feasible region; $g(27) = 0$
- $N(x) = \{x-1, x, x+1\}$



$L = f(x) + 0 \times max(g(x), 0)$

$(x = 27, \lambda = 0)$ is not a saddle point



$L = f(x) + 100 \times max(g(x), 0)$

$(x = 27, \lambda = 100)$ is a saddle point

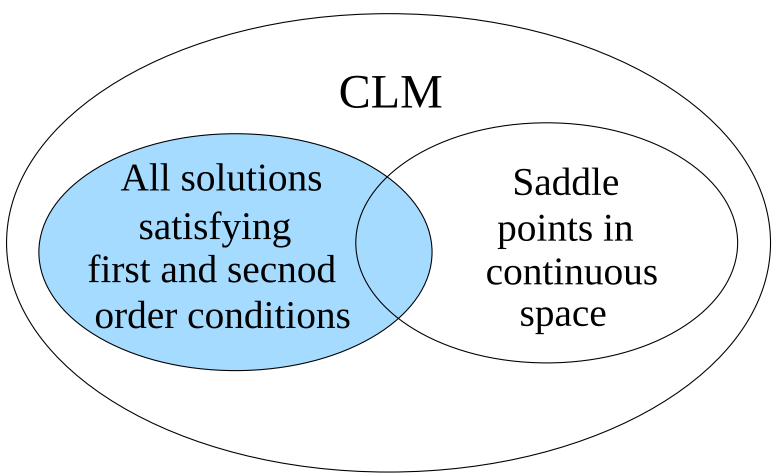
CLM, Saddle Points, and First-Order Conditions

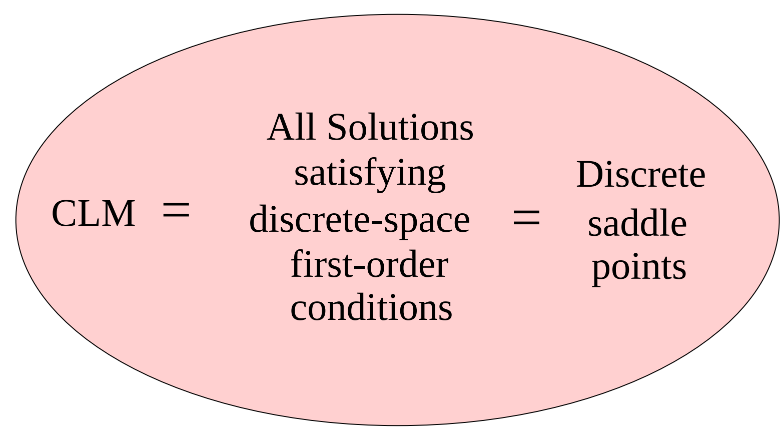
Continuous Space

- Saddle point $\Rightarrow$ CLM
- Set of saddle points $\neq$ set of points satisfying first- and second-order conditions
- Global optimization not meaningful

Discrete Space

- If $H(h(x))$ (transformations of $h(x)$) are all non-negative or non-positive, then set of saddle points $=$ set of CLM
- Set of saddle points $=$ set of solutions to first-order conditions
- Global optimization is meaningful





Proof: CLM $\Leftrightarrow$ Saddle Points

$\Leftarrow$: A saddle point is a CLM because it is feasible and is a local minimum

$\Rightarrow$: For CLM x^*, find λ^* to make $L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*)$ true or $f(x^*) \leq f(x) + \sum_i \lambda_i^* H(h_i(x))$

- Set initial $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*) = (0, \ldots, 0)$.
- Enumerate over all $x \in \mathcal{N}(x^*)$

  Case I: x is feasible, then any λ^* will be fine

  Case II: x is infeasible, (say $H(h_i(x)) > 0$) then we add enough penalty on that constraint function by setting

$$\lambda_i^* \leftarrow \max\left(\lambda_i^*, \frac{f(x^*) - f(x)}{H(h_i(x))}\right)$$

- Prove that (x^*, λ^*) is a saddle point

Proof: Saddle Points $\Leftrightarrow$ Solutions to First-Order Conditions

- A saddle point (x^*, λ^*): $L_d(x^*, \lambda) \le L_d(x^*, \lambda^*) \le L_d(x, \lambda^*)$

- First-order conditions:

$$\Delta L_d(x, \lambda) = 0, \quad h(x) = 0$$

$\Rightarrow$ A saddle point must satisfy first-order conditions because x^* is feasible and $L_d(x^*, \lambda^*) \le L_d(x, \lambda^*)$

$\Leftarrow$ A solution to first-order conditions must be a saddle point because it is feasible and is a local minimum

Transformations and Handling Inequality Constraint Functions

- Inequality constraints: $g(x) \leq 0 \Rightarrow \max(g(x), 0) = 0$
- Equality constraints: Transform $h(x)$ into $H(h(x))$
  - $H(h(x)) = 0 \Leftrightarrow h(x) = 0$
  - $H(h(x)) \geq 0$
- Examples
  - $H(h(x)) = h^2(x)$
  - $H(h(x)) = |h(x)|$
- Similar approach does not work in continuous space
  - Resulting function is not differentiable at $h(x) = 0$

Outline

- Previous work on nonlinear discrete constrained optimization
- Theory of discrete Lagrange-multiplier optimization
  - Basic concepts
  - Discrete-space first-order necessary and sufficient conditions
- Extension of theory to continuous and mixed-integer constrained optimization
- Seach algorithms
  - DLM: an implementation of the first-order search method
  - CSA: constrained simulated annealing with global convergence

Existing Work on Mixed-Integer Nonlinear Constrained Opt.

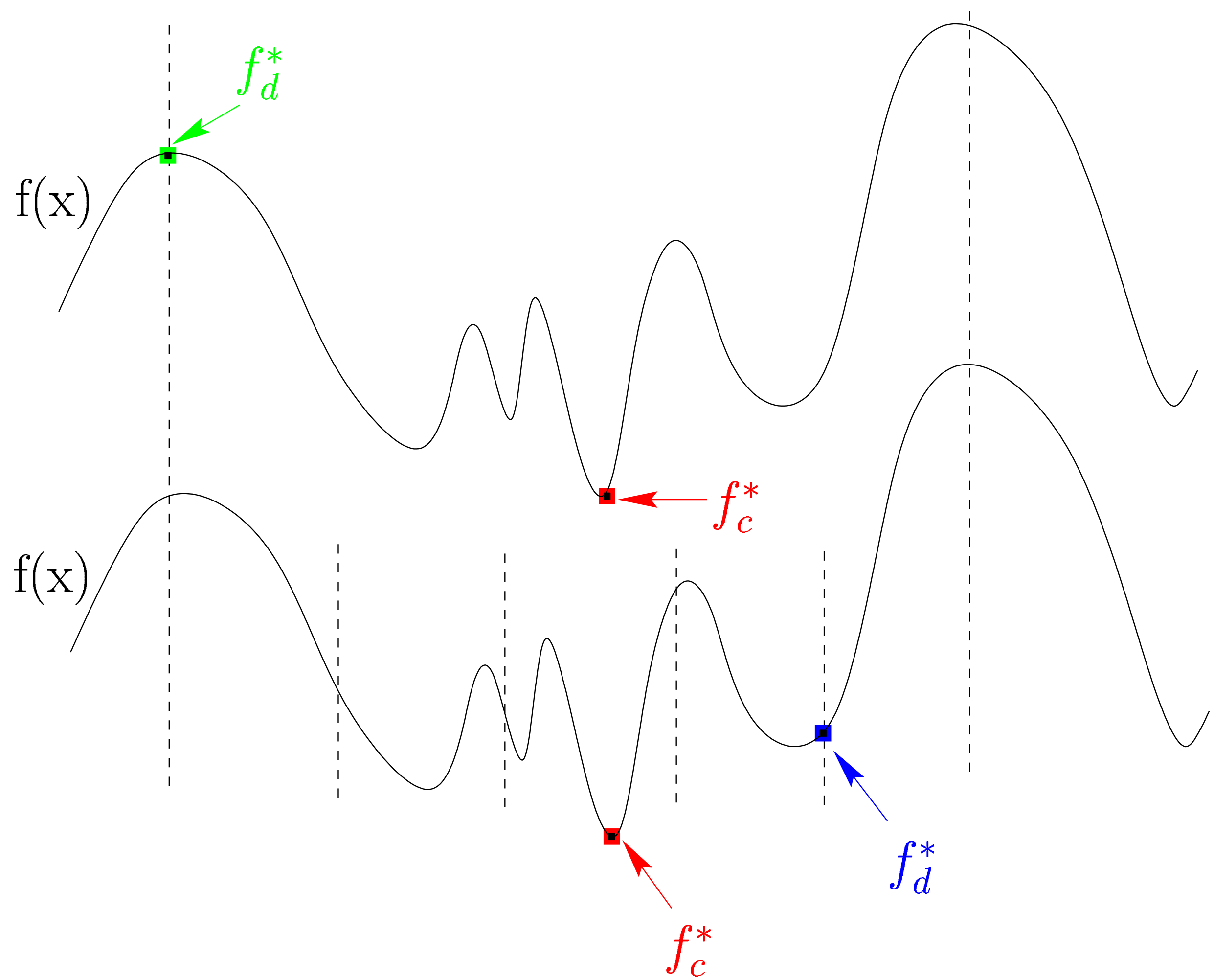
- Utilizing convexity of functions
  - Generalized Benders decomposition
  - Outer approximation
  - Generalized cross decomposition
  - Issue: convexity requirements are limited

- Penalty based methods
  - GA: apply non-uniform mutations to floating-point numbers
  - SA: generate candidate points on discrete grids
  - Issue: suitable static weights are hard to choose

Existing Work on Discretization

- Applying discretization methods to solve continuous problems
  - Solving differential equations
    - Divergence if step size not chosen correctly
  - Discretization in Semi-Infinite Programming (SIP) to make number of constraints finite
    - An ad hoc numerical method
  - Discretization of time progress in chemical engineering

Quality Improves as Discretization Level Increases



Extensions of Theory to Continuous/Mixed Space

- Theoretical foundation
  - Smoothness assumption - widely used Lipschitzian condition

$$|f(x_1) - f(x_2)| \leq l|x_1 - x_2|$$

  - Feasibility assumption - a constraint is satisfied if

$$|violation| \leq \theta$$

  - Given minimum grid size, s_{grid}, if $s_{grid} \leq \frac{\theta}{l}$, then the discrete optimal function value satisfies

$$\boxed{|f_c^* - f_d^*| \leq l \cdot s_{grid}}$$

    (f_c^* is the true continuous optimal function value)

- Implementation on digital computers
  - $s_{grid} = 10^{-15}$ (double precision), $\theta = 10^{-7}$, $|f_c^* - f_d^*| \leq 10^{-7}$, $l \leq 10^8$

Comparisons of Theory of Lagrange Multipliers

Continuous Space

- Theory based on regular points and differentiability
- First order necessary and second-order sufficient conditions
- No extensions to mixed-integer and discrete space
- Global optimization based on first-order necessary and second-order sufficient conditions is not meaningful
- No systematic way to find saddle points

Discrete Space

- Theory based on $N(x)$, $DMPD$, and discrete saddle points
- First-order necessary and sufficient conditions
- Theory can be extended to mixed-integer and continuous space
- Meaningful global optimization based on finding saddle points with optimal objective value
- First order search method for finding saddle points

Outline

- Previous work on nonlinear discrete constrained optimization
- Theory of discrete Lagrange-multiplier optimization
  - Basic concepts
  - Discrete-space first-order necessary and sufficient conditions
- Extension of theory to continuous and mixed-integer constrained optimization
- Search algorithms
  - DLM: an implementation of the first-order search method
  - CSA: constrained simulated annealing with global convergence

First-Order Local-Search Method

Continuous Space

- $x^{k+1} = x^k - \alpha_k \nabla_x L(x^k, \lambda^k)$
  $\lambda^{k+1} = \lambda^k + \alpha_k \nabla_\lambda L(x^k, \lambda^k)$
- Closed-form gradient functions (may be very slow or unstable due to function stiffness)
- Convergence time not bounded
- Solution checked using second-order conditions when search converges
- More efficient methods like SQP

Discrete Space

- $x^{k+1} = x^k \oplus \Delta_x L_d(x^k, \lambda^k)$
  $\lambda^{k+1} = \lambda^k + c_1 H(h(x^k))$
- Does not require gradient functions
- Convergence time not bounded
- Fixed-point theorem: a saddle point is reached iff procedure terminates

Experimental Results on DLM: MAX-SAT

- Test problems: 44 instances derived from DIMACS SAT benchmark
- Comparison with GRASP (a randomized adaptive search method)
- 50 times faster

Problem	# Clauses	GRASP (seconds)	DLM (PPro200)
jnh1-19	850	799.8	0.09
jnh201-220	800	692.9	0.01
jnh301-310	900	937.8	0.07

- Better quality of solutions

	Number of instances solved optimally
GRASP	3
DLM	44

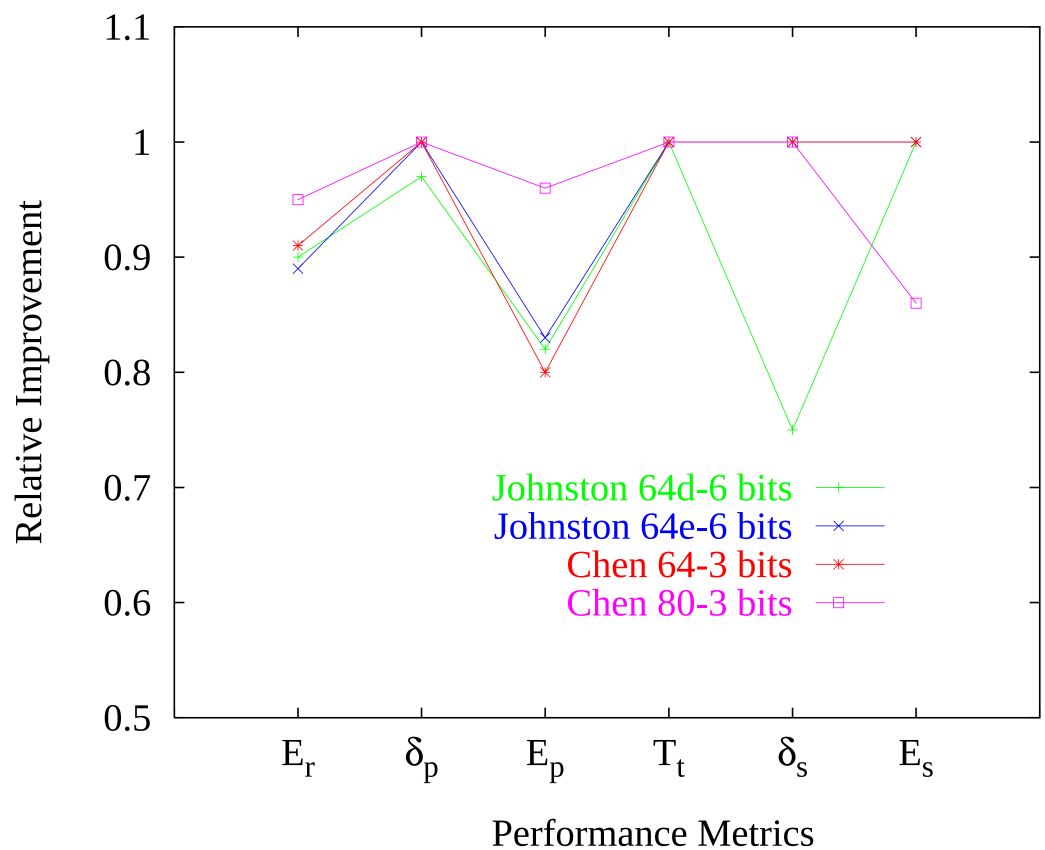
Experimental Results on DLM: SAT Problems

- 152 SAT benchmark instances in the DIMACS archive
- DLM can solve some hard-to-solve problems:
  - Hanoi4, Hanoi4-simple, par16-[1-5], par16-[1-5]-c, par32-[1-4]-c, f2000
  - Unresolved DIMACS benchmarks: Hanoi5, par32-5-c, par32-[1-5]

Problem Id	No. of Vars.	No. of Clauses	Succ. Ratio	CPU Time (P2/400 sec)	No. of Flips
hanoi4	718	4934	10/10	17765	$5.4 \cdot 10^8$
hanoi4$_s$	541	3008	10/10	17152	$9.2 \cdot 10^8$
par32-1-c	1315	5254	1/20	10388	$8.9 \cdot 10^8$
par32-2-c	1303	5206	1/20	123603	$9.2 \cdot 10^9$
par32-3-c	1325	5294	1/20	186274	$1.4 \cdot 10^{10}$
par32-4-c	1333	5326	1/20	139715	$1.1 \cdot 10^{10}$

Results: Multiplierless QMF Filter-Bank Design

- Powers-of-two (PO2) filter coefficient representation, e.g. $2^{-2} - 2^{-5}$
- Improvements over existing methods

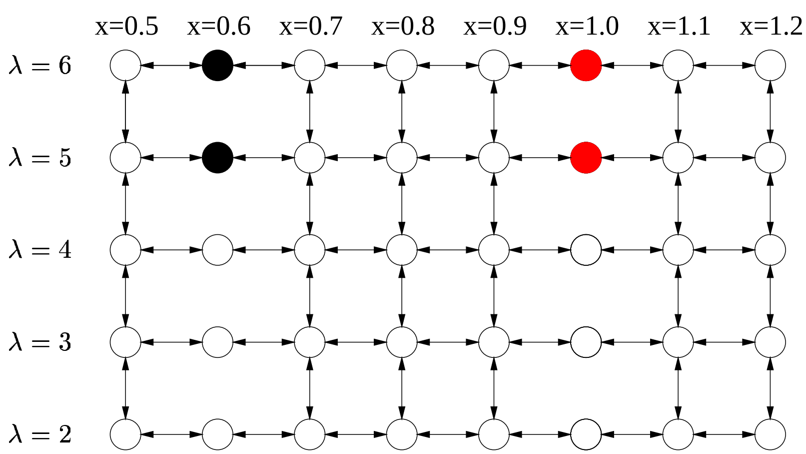


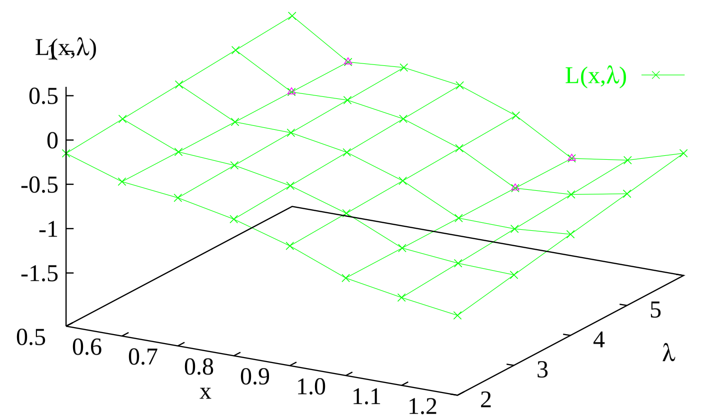
Constrained Simulated Annealing

- New assumptions
  - Bounded discrete variable space
  - Reachability of every state from every other state
- Search for discrete-space saddle points with minimum objective value
  - Descents in original-variable space
    - Smaller Lagrangian function value: accept with probability one
    - Larger Lagrangian function value: accept with temperature-dependent Metropolis probability
  - Ascents in Lagrange-multiplier space
    - Larger Lagrangian function value: accept with probability one
    - Smaller Lagrangian function value: accept with temperature-dependent Metropolis probability
- Logarithmically decreasing temperature schedule
- Asymptotic convergence to saddle point of minimum objective value with probability one (must be one of the CLMs regardless of $N(x)$)

Illustrative Example 1

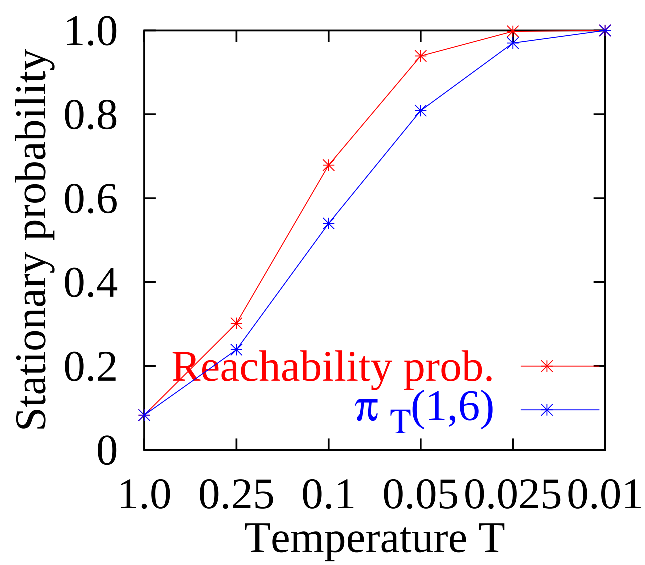
$$\begin{aligned} &\text{minimize} \quad f(x) = -x^2 \\ &\text{subject to} \quad h(x) = |(x-0.6)(x-1.0)| = 0 \end{aligned}$$



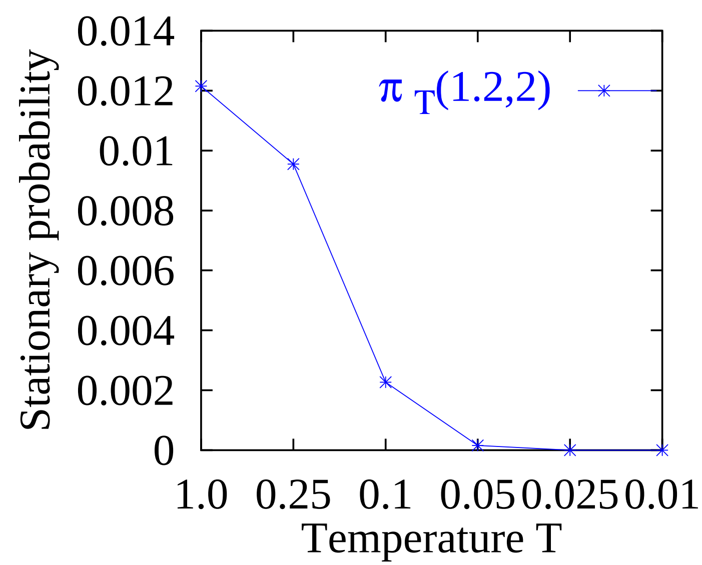


Illustrative Example 1: Convergence and Reachability Probabilities

Constrained global minimum



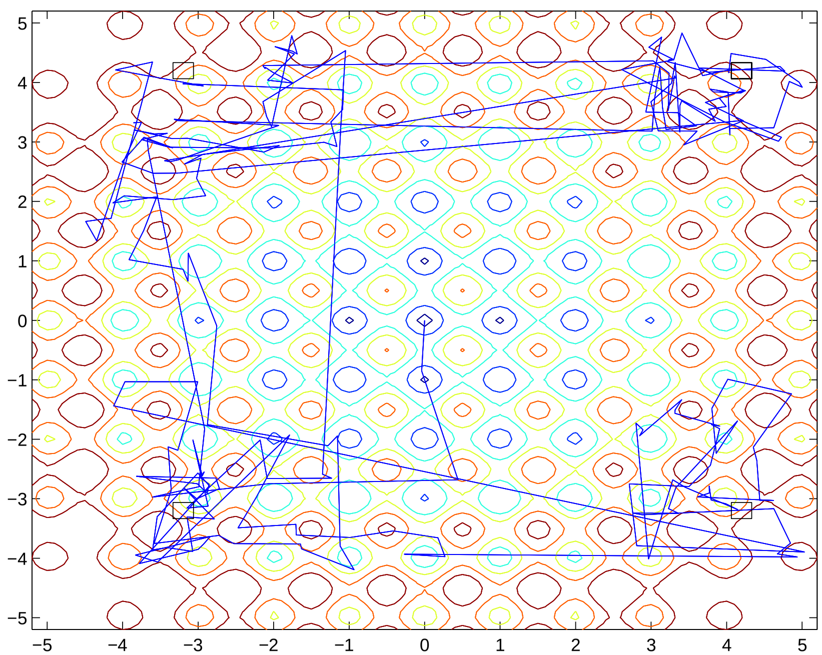
Minimum Lagrangian value



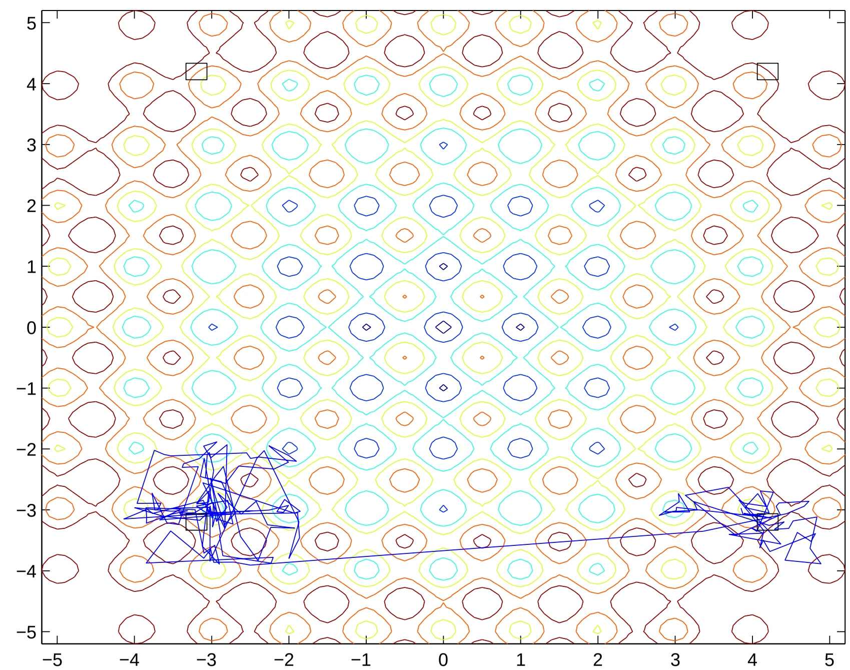
Illustrative Example 2

$$\text{minimize} \quad f(x) = 10n + \sum_{i=1}^{n}(x_i^2 - 10cos(2\pi x_i))$$

$$\text{subject to} \quad |(x_i - 4.2)(x_i + 3.2)| \leq 0 \qquad \text{where } n = 2,$$

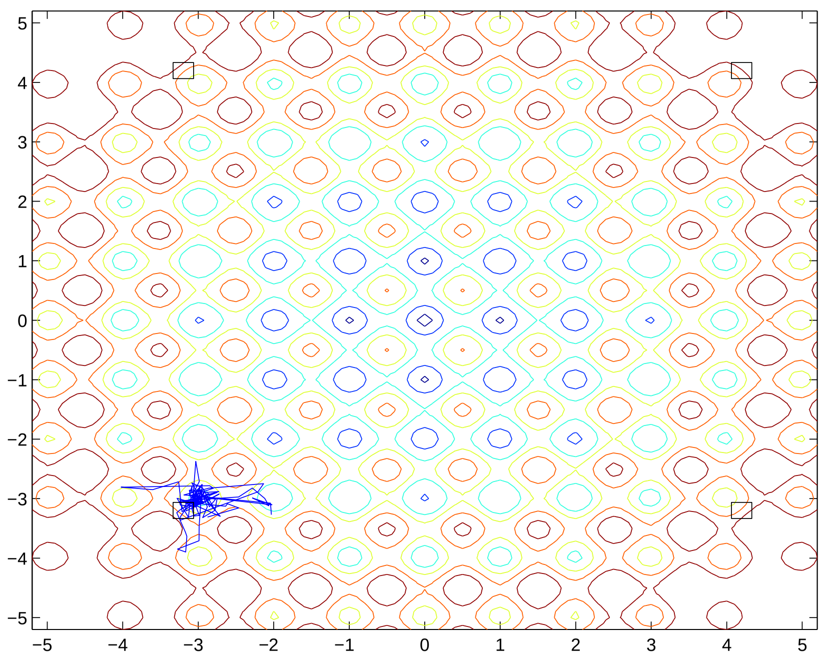


(a) T = 20

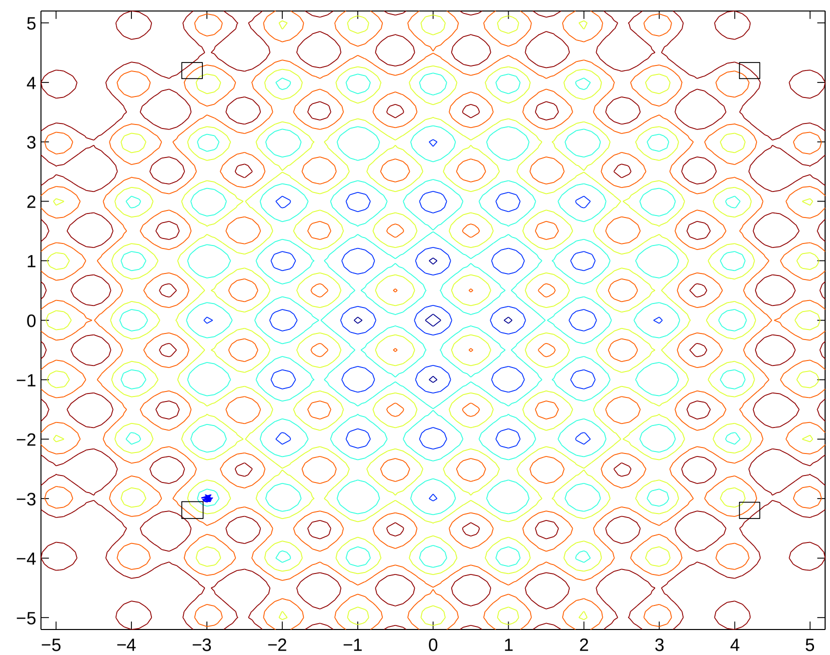


(b) T = 10.24

Illustrative Example 2 (cont'd)



(c) T = 8.192



(d) T = 0.45

Proof: Asymptotic Convergence of CSA

- Set N_T to be N_L (diameter of Markov chain modeling process)
- Markov chain is strongly ergodic if $\{T_k\}$ satisfies certain property
  - Markov chain has a unique stationary disribution
- Markov chain fits into framework of generalized SA (GSA)
  - Communication cost over each edge as function of L
  - Virtual energy over minimum spanning tree
- Markov chain modeling CSA converges to global minimum with probability 1
  - Minimum virtual energy at constrained global minima

Results: Nonlinear Continuous Optimization Problems

Problem ID	Global Solution	Original Continuous Versions						
		EA		SQP: DONLP2		CSA (average of 20 runs)		
		best solution	specific method	best solution	% with best sol.	best solution	% with best sol.	Time per run
G1 (min)	-15	-15	Genocop	-15	14.8%	-15	100%	17.04
G2 (max)	unknown	0.803553	S.T.	0.640329	0.4%	0.803619	100%	94.03
G3 (max)	1.0	0.999866	S.T.	1.0	93.1%	1.0	100%	69.62
G4 (min)	-30665.5	-30664.5	H.M.	-30665.5	61.4%	-30665.5	100%	2.70
G5 (min)	unknown	5126.498	D.P.	4221.956	94.0%	4221.956	100%	3.81
G6 (min)	-6961.81	-6961.81	Genocop	-6961.81	87.2%	-6961.81	100%	0.978
G7 (min)	24.3062	24.62	H.M.	24.3062	99.3%	24.3062	100%	14.71
G8 (max)	unknown	0.095825	H.M.	0.095825	44.9%	0.095825	100%	1.22
G9 (min)	680.63	680.64	Genocop	680.63	99.7%	680.63	100%	5.61
G10 (min)	7049.33	7147.9	H.M.	7049.33	29.1%	7049.33	100%	10.81

Objectives: linear, quadratic, polynomial, nonlinear

Constraints: linear inequalities, nonlinear equalities, nonlinear inequalities

No. of variables ≤ 20, No. of constraints ≤ 42

Results: Nonlinear Discretized and Mixed Problems

Problem ID	Global Solution	Discretized Versions			Mixed Versions		
		20 runs of CSA			20 runs of CSA		
		best solution	% with best sol.	Time per run	best solution	% with best sol.	Time per run
G1 (min)	-15	-15	100%	17.97	-15	100%	21.56
G2 (max)	unknown	0.803619	90%	99.03	0.803619	100%	100.62
G3 (max)	1.0	1.0	100%	70.15	1.0	100%	74.78
G4 (min)	-30665.5	-30665.5	100%	2.43	-30665.5	100%	3.04
G5 (min)	unknown	4221.956	90%	3.32	4221.956	100%	4.16
G6 (min)	-6961.81	-6961.81	90%	0.962	-6961.81	100%	1.00
G7 (min)	24.3062	24.3062	100%	17.48	24.3062	95%	17.39
G8 (max)	unknown	0.095825	100%	1.37	0.095825	100%	1.33
G9 (min)	680.63	680.63	100%	6.83	680.63	100%	6.76
G10 (min)	7049.33	7049.33	100%	13.31	7049.33	100%	12.57

Each unit of discretized variable is discretized into 10^5 intervals

Variables are offset so that the best solutions in discrete/mixed cases coincides with those in continuous cases

Summary

- Formal mathematical foundation
  - Discrete neighborhoods, DMPD, and first-order necessary and sufficient conditions
  - Extensions to nonlinear continuous and mixed optimization problems
- Generality of theory to handle discrete nonlinear constrained problems
  - Does not require closed-form objective/constraint functions
  - Does not require closed-form gradient functions
- Efficient algorithms
  - Efficient local search algorithms
  - Global optimization algorithms with asymptotic convergence