Stochastic Global Optimization Techniques

Electromagnetic Method

Outline

PART I

- Global Optimization Problem
  - Definition and challenges.
- Proposed Approach
  - General scheme and basic procedures.
- Computational Experience
  - Experimentation with local procedure and comparison with other methods.

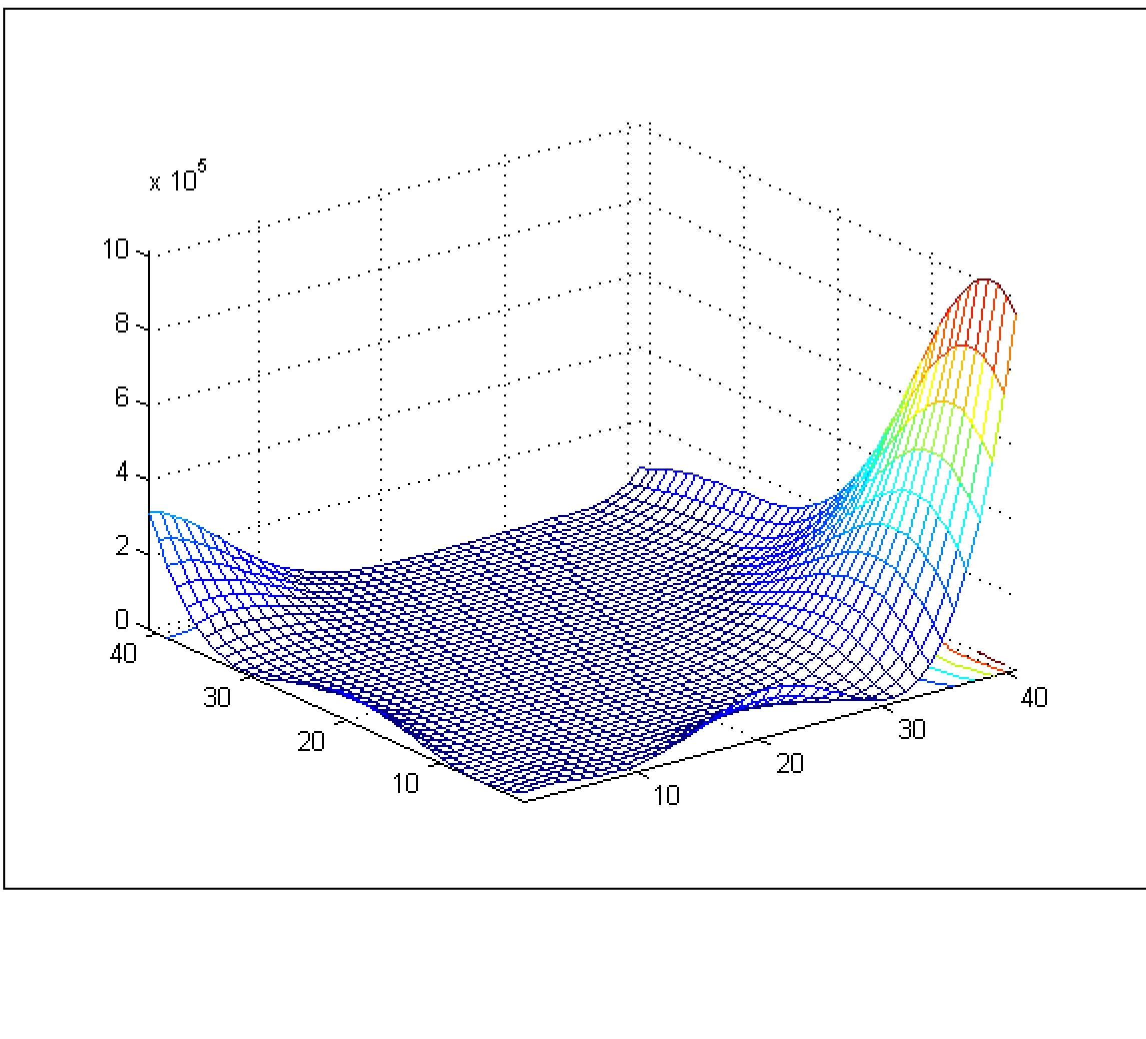
PART II

- Mathematical Theory
  - Refined algorithm and the convergence theorem.
- Future Research Directions

Global Optimization Problem

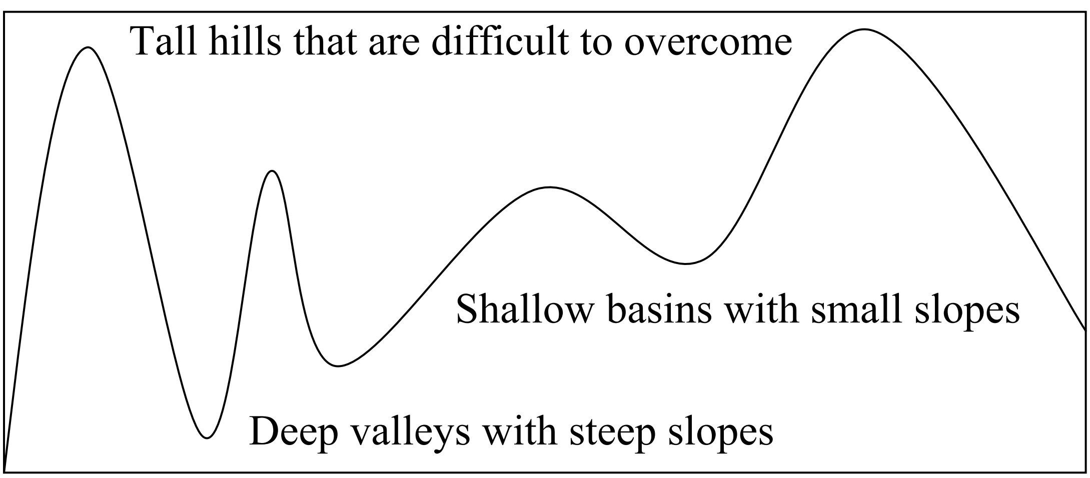
$$\min f(x)$$
$$s.t.$$
$$x \in [l,u]$$

where $[l,u] := \{x \in \Re^n \mid l_k \leq x_k \leq u_k, k = 1,\ldots,n\}$



Challenges

- Local v.s. global minimum
- nonlinear functions
- non-differentiable functions
- multi-modality of the functions
- combinatorial structure
- extensive function evaluations

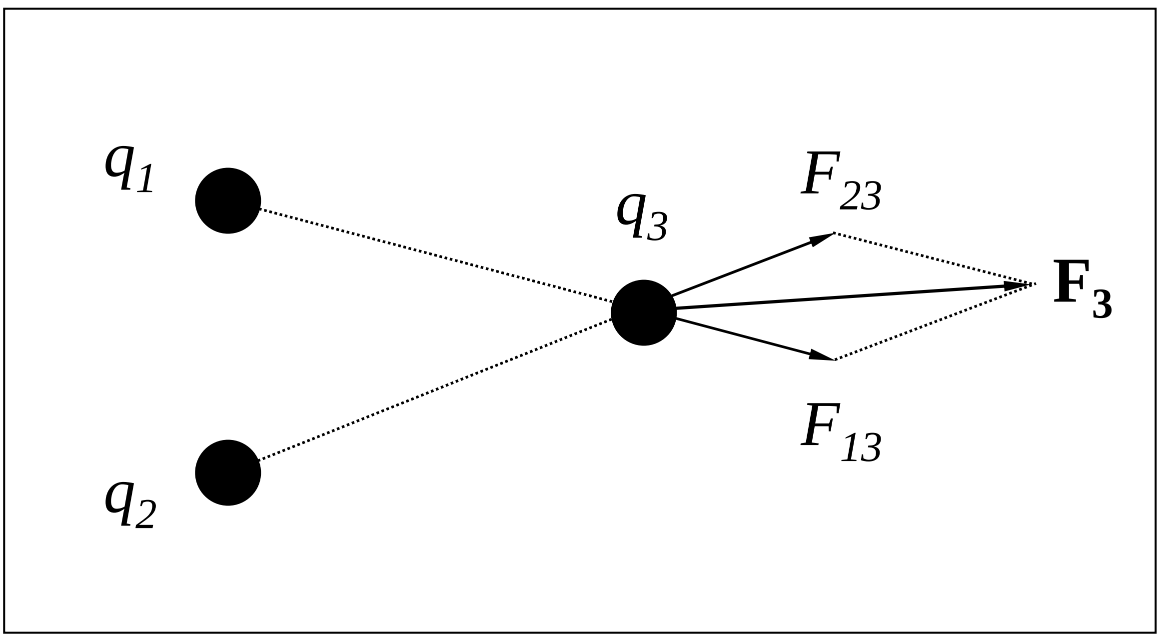


Available Approaches

- simulated annealing
- two-phase methods
- evolutionary methods
- partitioning methods

Principle of Superposition

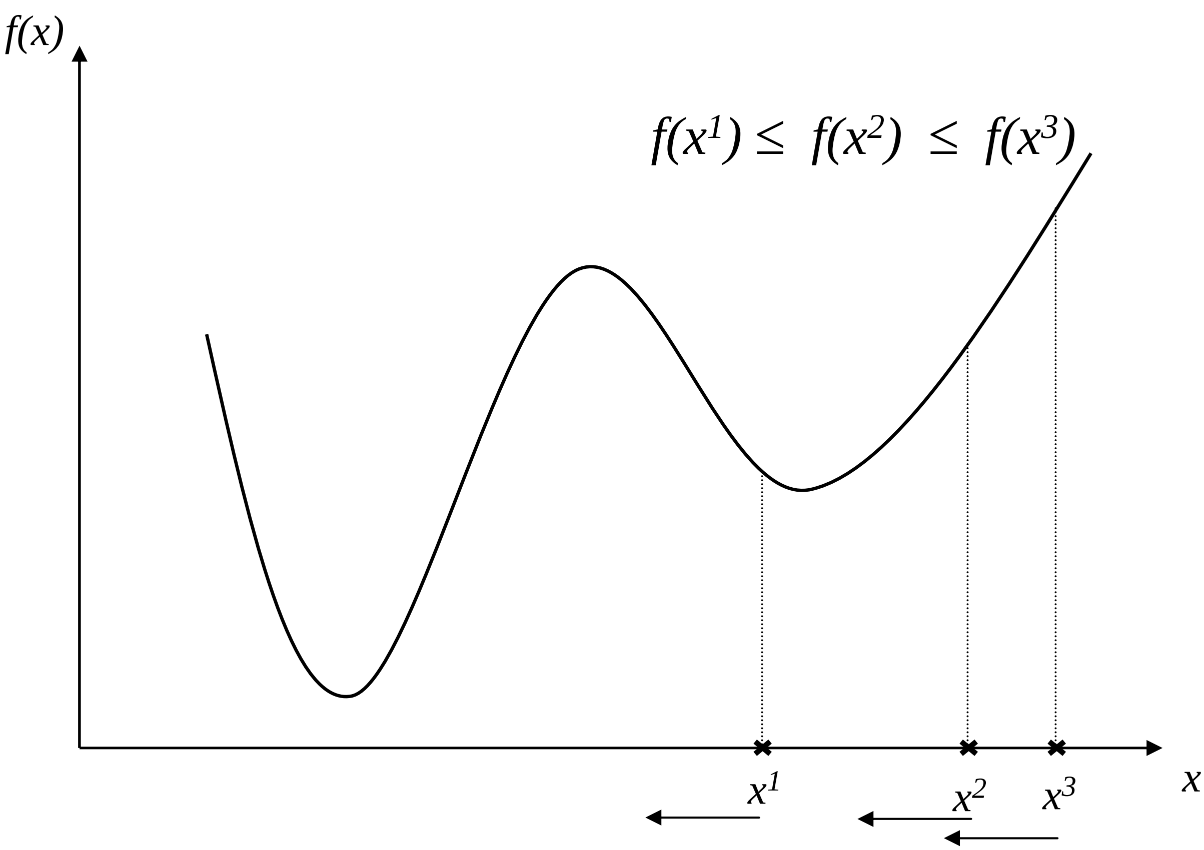
- Attraction and Repulsion
- *Principle of Superposition*



$$\vec{F}_{i3} = \frac{q_3 q_i}{4\pi\varepsilon_0\varepsilon_r r^2}\vec{e_r} \quad , i = 1,2$$

Proposed Approach (EM)

- *m* points are randomly generated from the feasible region.
- Points with better function values attract other points.
- Points with worse objective values repel other points.
- After moving points to new locations the objective function values are updated.



General Scheme

- EM (m, MAXITER, LSITER,δ)

Initialize()

{m points are randomly generated from the feasible domain}

iteration ← 1

while iteration < MAXITER **do**

{*Local(LSITER, δ)*}

F ← CalcF()

{The total force vector on each point is calculated.}

Move(F)

{The points are moved in the direction of the force, F.}

iteration ← iteration + 1

end while

Total Force Calculation

The charge of each point, q^i, is a nonnegative value, which determines the power of attraction or repulsion for point, i.

$$q^i = \exp\left(-n \frac{f(x^i) - f(x^{best})}{\sum_k \left(f(x^k) - f(x^{best})\right)}\right), \; i = 1,2,\ldots,m$$

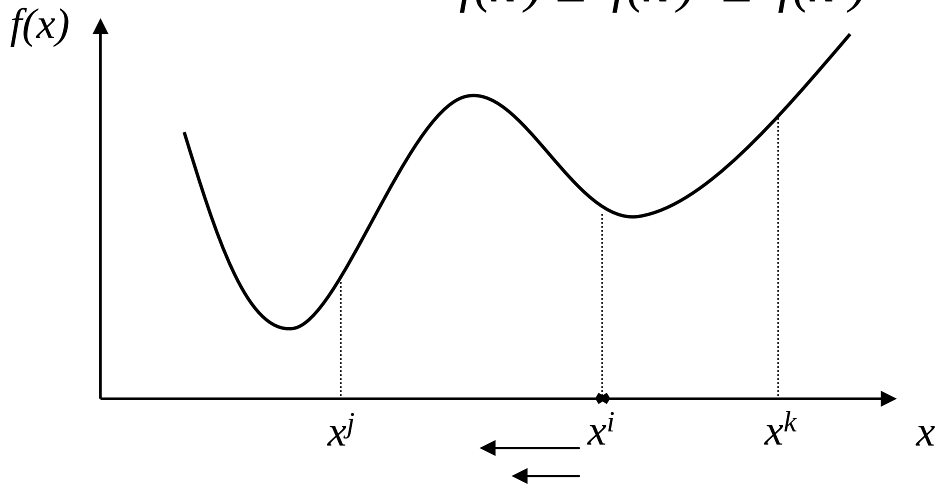
After determining the charge of each point, the total force, F^i, exerted on point i is computed,

$$F^i = \sum_{j \neq i} (x^j - x^i) \frac{q^i q^j}{\left\| x^j - x^i \right\|^2}, \; i = 1,2,\ldots,m$$

Total Force Calculation

In the algorithm:

$f(x^j) \leq f(x^i) \leq f(x^k)$



$F^i = F_{ji} + F_{ki}$

{Attraction}

$f(x^j) < f(x^i)$

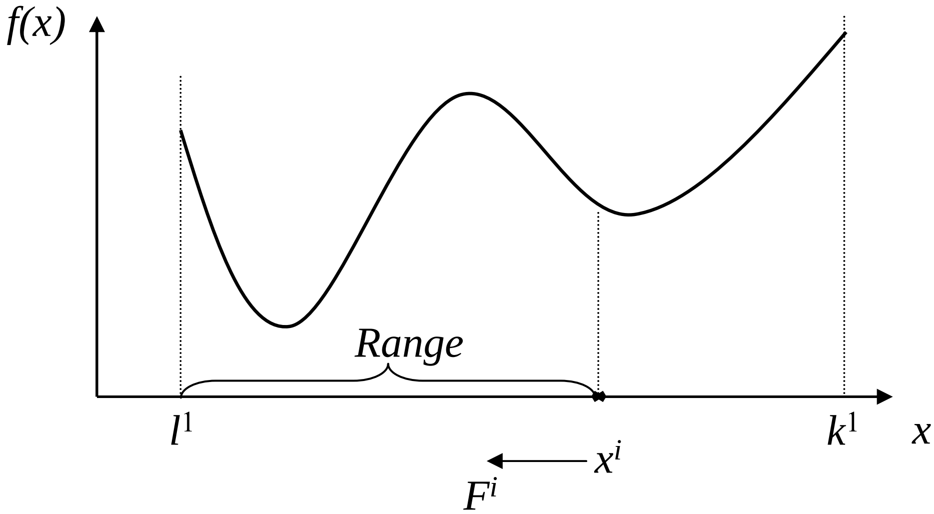
$$F^i \leftarrow F^i + (x^j - x^i)\frac{q^i q^j}{\left\| x^j - x^i \right\|^2}$$

{Repulsion}

$f(x^k) > f(x^i)$

$$F^i \leftarrow F^i - (x^k - x^i)\frac{q^i q^k}{\left\| x^k - x^i \right\|^2}$$

Movement



After evaluating the total force on point i, F^i, the point is moved.

$$x^i = x^i + \lambda \frac{F^i}{\left\|F^i\right\|}(Range)$$

λ: random number, U~(0,1).

Range : allowed feasible movement.

Computational Experience

- 15 frequently cited functions are selected.
- 25 runs are taken for each function.

The following input parameters are used:

Function Name	n	m	MAXITER	LSITER	δ
Complex	2	10	50	10	5.0e-3
Davis	2	20	50	30	5.0e-3
Epistacity(4)	4	30	50	10	1.0e-3
Epistacity(5)	5	40	100	20	1.0e-3
Griewank	2	30	100	20	1.0e-3
Himmelblau	2	10	50	5	1.0e-3
Kearfott	4	10	50	5	1.0e-3
Levy	10	20	75	5	1.0e-3
Rastrigin	2	20	50	10	5.0e-3
Sine Envelope	2	20	75	10	5.0e-4
Stenger	2	10	75	10	1.0e-3
Step	5	10	50	5	1.0e-3
Spiky	2	30	75	10	1.0e-3
Trid(5)	5	10	125	50	1.0e-2
Trid(20)	20	40	500	150	1.0e-3

Computational Experience

Results of EM for the general test functions.

Function Name	Avg. Evals.	Avg $f(x)$	Best $f(x)$	Known Optimum
Complex	363	0.0175	0.0158	0.0
Davis	622	1.6157	1.5641	0.0
Epistacity(4)	1079	0.0379	0.0149	0.0
Epistacity(5)	2603	0.0355	0.0207	0.0
Griewank	1914	0.0896	0.0032	0.0
Himmelblau	84	0.0934	0.0759	0.0
Kearfott	231	0.0008	0.0000	0.0
Levy	835	0.1429	0.0303	0.0
Rastrigin	141	-1.9566	-1.9871	-2.0
Sine Envelope	962	0.0744	0.0400	0.0
Stenger	282	0.0020	0.0019	0.0
Step	728	0.0000	0.0000	0.0
Spiky	1702	-38.6378	-38.7251	-38.85
Trid(5)	968	-28.2997	-29.0324	-30.0
Trid(20)	43354	-33.2567	-177.6124	-1520.0

EVAL. : average number of function evaluations

AVG. f(x) : average objective function value in 25 runs

BEST f(x) : best objective function value in 25 runs

KNOWN OPT. : known optimum value of the function

Computational Experience

Results of EM with Local procedure applied to all points

Function Name	Avg. Evals.	Avg $f(x)$	Best $f(x)$	Known Optimum
Complex	5534	0.0000	0.0000	0.0
Davis	8091	0.4088	0.1322	0.0
Epistacity(4)	32823	0.0003	0.0001	0.0
Epistacity(5)	189769	0.0001	0.0001	0.0
Griewank	50790	0.0000	0.0000	0.0
Himmelblau	3287	0.0000	0.0000	0.0
Kearfott	6190	0.0000	0.0000	0.0
Levy	44814	0.0001	0.0000	0.0
Rastrigin	10359	-2.0000	-2.0000	-2.0
Sine Envelope	15629	0.0116	0.0001	0.0
Stenger	8316	0.0000	0.0000	0.0
Step	5743	0.0000	0.0000	0.0
Spiky	10264	-38.8004	-38.8492	-38.85
Trid(5)	37154	-29.9979	-29.9999	-30.0
Trid(20)	~1.5e+6	-1519.6117	-1519.9768	-1520.0

EVAL. : average number of function evaluations

AVG. f(x) : average objective function value in 25 runs

BEST f(x) : best objective function value in 25 runs

KNOWN OPT. : known optimum value of the function

Computational Experience

Results of EM with Local procedure applied to x^{best} only.

Function Name	Avg. Evals.	Avg $f(x)$	Best $f(x)$	Known Optimum
Complex	598	0.0000	0.0000	0.0
Davis	832	0.4538	0.2356	0.0
Epistacity(4)	1580	0.0002	0.0001	0.0
Epistacity(5)	4123	0.0002	0.0000	0.0
Griewank	2470	0.0000	0.0000	0.0
Himmelblau	520	0.0001	0.0000	0.0
Kearfott	712	0.0000	0.0000	0.0
Levy	2783	0.0001	0.0000	0.0
Rastrigin	792	-1.9898	-2.0000	-2.0
Sine Envelope	1007	0.0352	0.0097	0.0
Stenger	724	0.0000	0.0000	0.0
Step	870	0.0000	0.0000	0.0
Spiky	1520	-38.6684	-38.8486	-38.85
Trid(5)	1870	-29.9963	-29.9997	-30.0
Trid(20)	99731	-1519.4472	-1519.5543	-1520.0

EVAL. : average number of function evaluations

AVG. f(x) : average objective function value in 25 runs

BEST f(x) : best objective function value in 25 runs

KNOWN OPT. : known optimum value of the function

Comparison of EM Versions

Comparison of EM versions in terms of number of function evaluations.

Function	**None**	**All**	**Best**
Complex	363	5534	598
Davis	622	8091	832
Epistacity(4)	1079	32823	1580
Epistacity(5)	2603	189769	4123
Griewank	1914	50790	2470
Himmelblau	84	3287	520
Kearfott	231	6190	712
Levy	835	44814	2783
Rastrigin	141	10359	792
Sine Envelope	962	15629	1007
Stenger	282	8316	724
Step	728	5743	870
Spiky	1702	10264	1520
Trid(5)	968	37154	1870
Trid(20)	43354	1.5e+6	99731

None: Local procedure is not applied.

All: Local procedure applied to all points

Best: Local procedure applied to current best point.

Comparison of EM with other methods

Comparison of EM with different methods in terms of number of function evaluations

Method	S5	S7	S10	H3	H6	GP	BR	C6	SHU
Bremmerman	(a)	(a)	(a)	(a)	(a)	(a)	250		
Mod. Bremmerman	(a)	(a)	(a)	(a)	515	300	160		
Zilinskas	(a)	(a)	(a)	8641			5129		
Gomulka-Branin	5500	5020	4860						
Törn	3679	3606	3874	2584	3447	2499	1558		
Gomulka-Törn	6654	6084	6144						
Gomulka-V.M.	7085	6684	7352	6766	11125	1495	1318		
Price	3800	4900	4400	2400	7600	2500	1800		
Fagiuoli	2514	2519	2518	513	2916	158	1600		
De Biase-Frontini	620	788	1160	732	807	378	587		
Mockus	1174	1279	1209	513	1232	362	189		
Bélisle et al. (b)				339	302	4728	1846		
Boender et al. (f)	567	624	755	235	462	398	235		
Snyman-Fatti (f)	845	799	920	365	517	474		178	
Kostrowicki-Piela (g)	(c)	(c)	(c)	200	200	120		120	
Yao (f)								1132	6000
Perttunen (f)	516	371	250	264		82	97	54	197
Perttunen-Stuckman (f)	109	109	109	140	175	113	109	96	(a)
Jones et al. (h)	155	145	145	199	571	191	195	285	2967
Storn-Price (d)	6400	6194	6251	476	7220	1018	1190	416	1371
MCS (e) (f)	83	129	103	79	111	81	41	42	69
EM	3368	1782	5620	1114	2341	420	315	233	358

(a) Method converged to a local minimum.
(b) Average number of function evaluations when converges. For H6, converged only 70 percent of time.
(c) Global minimum not found within 12000 function calls.
(d) Average over 25 cases. For H6, average over 24 cases only; one case did not converge within 12000 function calls.
(e) We selected the version that gives the best results.
(f) Recent methods that use first or second order information.
(g) Requires closed form for a particular integral.
(h) Partitions the search space into hyper-rectangles.

Missing entry means that no result is available from the literature.

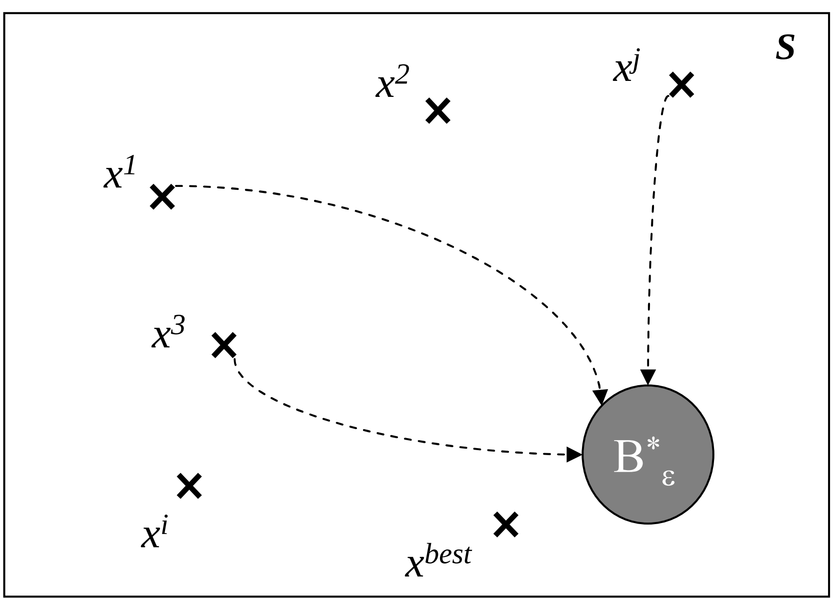
Intuitive Proof

Main Task:
Show that in the long run at least one of the *m* points of the population moves to the vicinity of a *global minimum* with probability one.

$x^* \in S \subset \Re^n$, is *global minimum* on *S* if $f(x^*) \leq f(x), \forall x \in S$

Set of *ε–optimal* points is denoted by

$$B_\varepsilon^* = \{x \in S : \left| f(x) - f(x^*) \right| \leq \varepsilon\}, \forall \varepsilon > 0$$



The algorithm converges when

$$\{x^1, x^2, \ldots, x^m\} \cap B^*_\varepsilon \neq \varnothing$$

Intuitive Proof

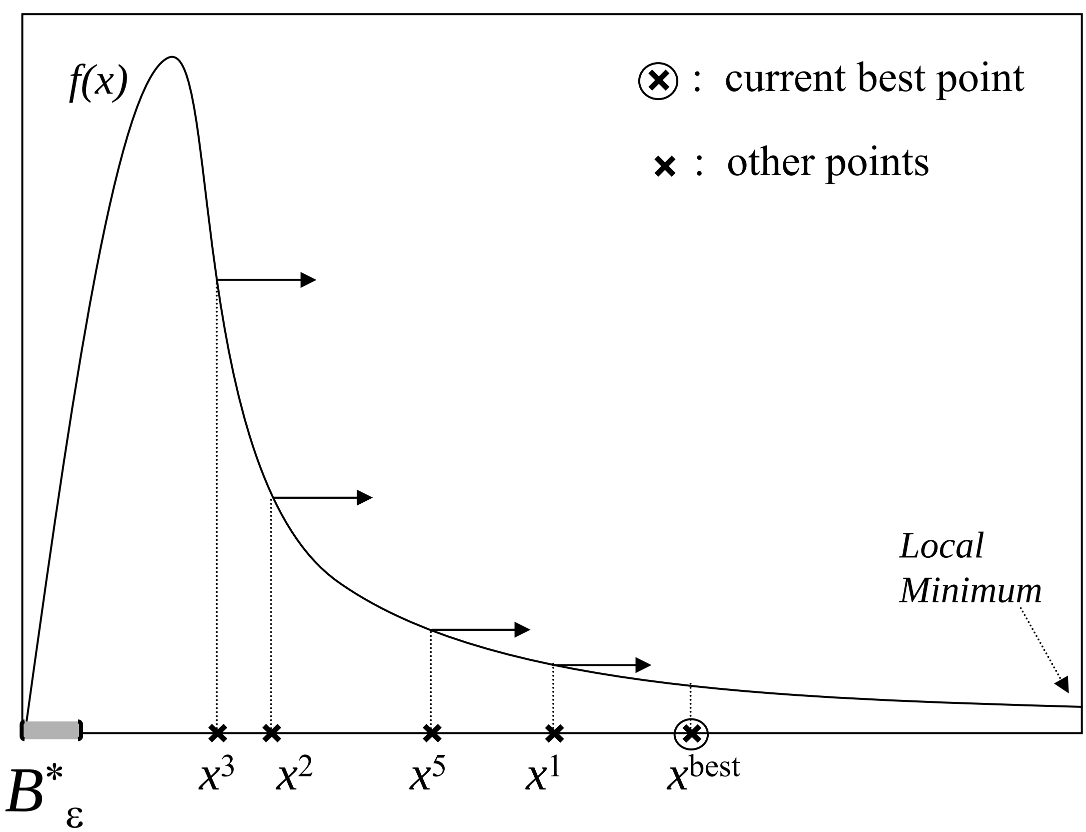
Basic requirement:
Ensure that at each iteration, k, there is a positive probability to move to any full-dimensional subset B of S.

ρ : probability of moving to subset B at each iteration.

$$1 - \lim_{k \uparrow \infty} (1-\rho)^k = 1$$

Assumption: B^*_ε has a nonzero measure. (non-degeneracy)

Premature Convergence



The Perturbed Point

A *perturbed point*, x^p, is introduced to avoid premature convergence.

The perturbed point is repelled (attracted) with certain probability although it has been attracted (repelled).

In the algorithm, the perturbed point is selected to be the farthest one from the current best point.

Assumption: (non-degeneracy)

$$\text{rank}\{x^1, x^2, \ldots, x^m\} = n$$

Nomenclature

$\mathbf{x} = \{x^1, x^2, \ldots, x^m\}$: a collection of *m* points forms a *state*.

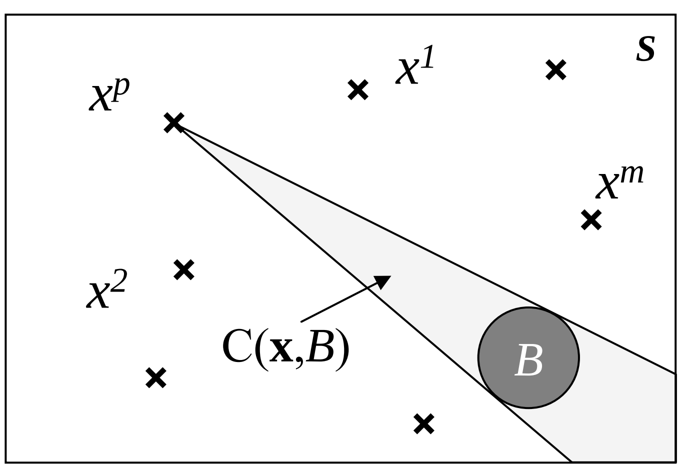
$P\{Y_{k+1} = \mathbf{y} \mid Y_k = \mathbf{x}\}$: the conditional probability of making a transition from state $\mathbf{x}$ to state $\mathbf{y}$.

In particular we will be interested in

$$P\{Y_{k+1} \cap B \neq \emptyset \mid Y_k = \mathbf{x}\}, B \subset S$$

In the proof we will concentrate on the perturbed point.

$C(\mathbf{x},B)$: the *truncated cone.*



Analysis

We try to show:

$$\lim_{k \uparrow \infty} P_{\mathbf{x}}\{Y_k \cap B_{\varepsilon}^{*} \neq \emptyset\} = 1, \forall \mathbf{x}$$

We need to ensure the following:

1. At *each iteration* of the algorithm, there is a positive probability to move to any positive measure set B in the feasible region S.

$$\rho(\mathbf{x},B) = P\{Y_{k+1} \cap B \neq \emptyset \mid Y_k = \mathbf{x}\} > 0, \forall \mathbf{x}$$

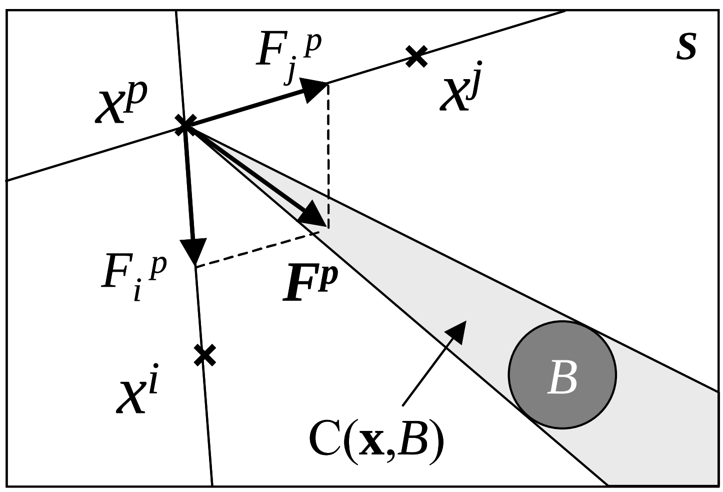
2. *In the limit* there is a nonzero lower bound for this probability.

$$\rho^{*} = \inf_{\mathbf{x},B} \{\rho(\mathbf{x},B)\} > 0$$

At Each Iteration

$$\rho(\mathbf{x},B) = P\{Y_{k+1} \cap B \neq \emptyset \mid Y_k = \mathbf{x}\} > 0, \forall \mathbf{x}$$

- It suffices to show that there is a positive probability for x^p to move into B. To do this we discuss the possibility of producing F^p that falls in C($\mathbf{x}$,B).



(F_j^p and F_i^p are component vectors)

- By the second assumption the collection of vectors, x^j-x^p, has full rank, so every direction around x^p can be generated.

- C($\mathbf{x}$,B) has a positive n dimensional volume. Thus, there is a positive probability of generating an F^p that lies in C($\mathbf{x}$,B).

In The Limit

$$\rho^* = \inf_{\mathbf{x}, B} \{\rho(\mathbf{x}, B)\} > 0$$

Problem : C($\mathbf{x}$,B) becomes degenerate.
(If the distance between x^p and a point in B becomes very large)

Solution: S is a compact set in $\Re^n$.

Problem: A component force dominates the others.
(If a point is too close to the perturbed point)

Solution: Any oversized component force is truncated in the refined algorithm.

Problem: A component force becomes zero.
(If a point is too far away from the perturbed point.)

Solution: Formally given in Lemma 1.

Lemmas

Lemma 1: For any direction (x^j-x^p) the magnitude of the component force F^j_p is nonzero for all $j=1,2,\ldots,m$ and $j \neq p$.

Lemma 2: $P\{Y_{k+1} \cap B^*_\varepsilon \neq \emptyset \mid Y_k = \mathbf{x}\} = 1$ when $Y_k \cap B^*_\varepsilon \neq \emptyset$

"Absorbing Event"

Convergence Theorem

Theorem: Provided that the assumptions given before hold, the EM algorithm converges to B^*_ε with probability one, i.e.,

$$\lim_{k \uparrow \infty} P_{\mathbf{x}}\{Y_k \cap B_\varepsilon^* \neq \emptyset\} = 1, \forall \mathbf{x} \in \mathrm{X}_m$$

Sketch of the proof:

$$\lim_{k \uparrow \infty} P_{\mathbf{x}}\{Y_k \cap B_\varepsilon^* \neq \emptyset\} = 1 - \lim_{k \uparrow \infty} P_{\mathbf{x}}\{Y_k \cap B_\varepsilon^* = \emptyset\}$$

$$\geq 1 - \lim_{k \uparrow \infty} (1-\rho^*)^k = 1$$

Computational Experience

Results for Dixon and Szegö (1978) Functions with the refined algorithm

Function	n	m	MAXITER	Avg. Evals.	Avg $\boldsymbol{f}(\boldsymbol{x})$	Best $\boldsymbol{f}(\boldsymbol{x})$	$\boldsymbol{f_{glob}}$
Shekel [S5]	4	40	150	2800	-9.54637	-10.1532	-10.1532
Shekel [S7]	4	40	150	1608	-10.4024	-10.4029	-10.4029
Shekel [S10]	4	40	150	5445	-10.5109	-10.5109	-10.5364
Hartman [H3]	3	30	75	1303	-3.8626	-3.8628	-3.8628
Hartman [H6]	6	30	75	2206	-3.3045	-3.3224	-3.3224
Goldstein Price [GP]	2	20	50	421	3.0001	3.0000	3.0000
Branin [BR]	2	20	50	393	0.3979	0.3979	0.3979
Six Hump Camel [C6]	2	20	50	253	-1.0316	-1.0316	-1.0316
Shubert [SHU]	2	20	50	265	-185.1975	-186.7309	-186.7309

Results with the previous algorithm

Function	n	m	MAXITER	Avg. Evals.	Avg $f(x)$	Best $f(x)$	f_{glob}
Shekel [S5]	4	40	150	3368	-9.7320	-10.1532	-10.1532
Shekel [S7]	4	40	150	1782	-10.4024	-10.4029	-10.4029
Shekel [S10]	4	40	150	5620	-10.5109	-10.5109	-10.5364
Hartman [H3]	3	30	75	1114	-3.8625	-3.8628	-3.8628
Hartman [H6]	6	30	75	2341	-3.3072	-3.3224	-3.3224
Goldstein Price [GP]	2	20	50	420	3.0001	3.0000	3.0000
Branin [BR]	2	20	50	315	0.3980	0.3979	0.3979
Six Hump Camel [C6]	2	20	50	233	-1.0316	-1.0316	-1.0316
Shubert [SHU]	2	20	50	358	-186.7227	-186.7309	-186.7309

Computational Experience

Results for Additional Test Set

Function	n	m	MAXITER	Avg. Evals.	Avg $\boldsymbol{f}(\boldsymbol{x})$	Best $\boldsymbol{f}(\boldsymbol{x})$	$\boldsymbol{f_{glob}}$
Rosenbrock	2	25	75	2117	0.0022	0.0000	0.0000
Shekel's Foxholes	2	50	100	4001	0.9980	0.9980	0.9980
Sphere	3	10	25	492	0.0000	0.0000	0.0000
Neumaier II	4	50	250	12863	0.0192	0.0000	0.0000
Neumaier III	10	50	150	11211	-209.7133	-209.9176	-210.00
Hyper-ellipsoid	30	50	50	6162	0.0000	0.0000	0.0000
Ackley	30	50	300	34753	0.9718	0.4018	0.0000

Future Research Directions

- Other Test Functions
  - Some real world problems with high dimensionality.
- Expected Number of Iterations
  - The expected hitting time to solve a given problem.
- Handling Complex Constraints
  - Set of equality and inequality constraints.
  - Initial step, a penalty method.
- Discrete Optimization Problems
  - Relaxing integrality constraints.
  - Designing the discrete version of the algorithm.
- Merging with Other Methods - Using First or Second Order Information
  - Effective methods for local refinement.