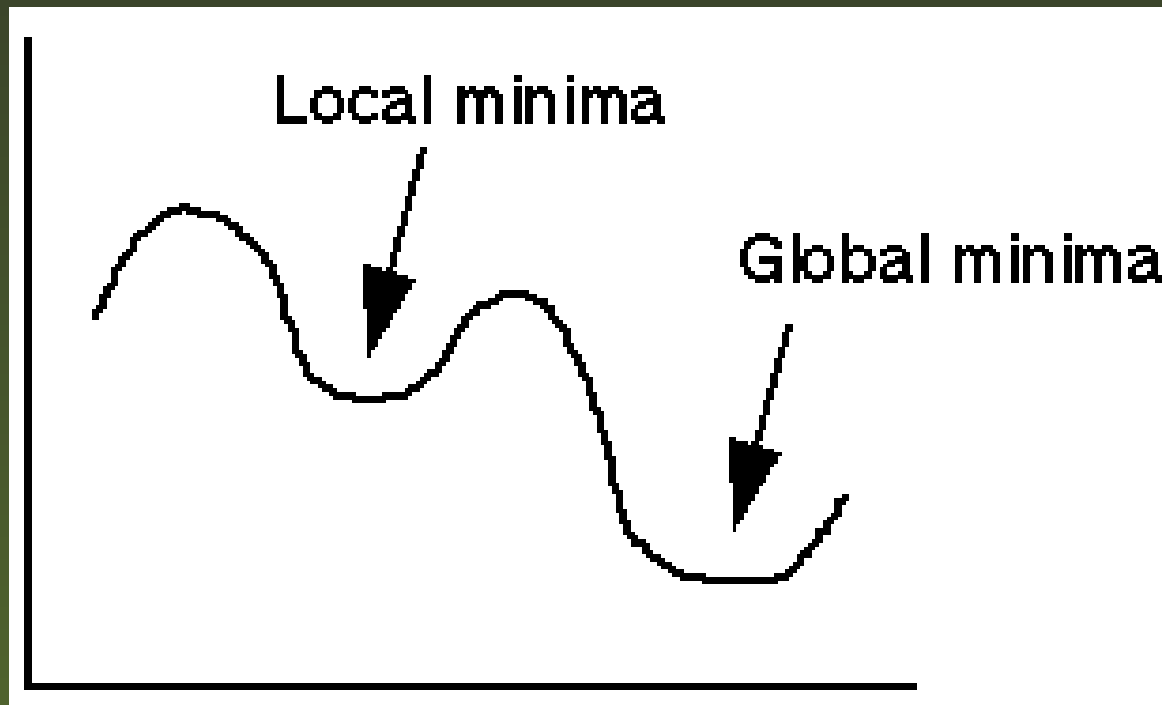# From Local to Global in Numerical Optimization

Panos M. Pardalos

Center for Applied Optimization, ISE & CISE Departments

Biomedical Engineering Department, McKnight Brain Institute

University of Florida, Gainesville, FL 32611

http://www.ise.ufl.edu/pardalos/

1 Global Continuous (or Discrete) Optimization Problem

$$f^* = f(x^*) = global\ min_{x \in D} f(x)\ (or\ max_{x \in D} f(x))$$

1 Global Continuous (or Discrete) Optimization Problem

References

[1] R. HORST AND P. M. PARDALOS (Editors), *Handbook of Global Optimization*, Kluwer Academic Publishers, 1995.

[2] P. M. PARDALOS AND H. E. ROMEIJN (Editors), *Handbook of Global Optimization Vol. 2*, Kluwer Academic Publishers, 2002.

[3] R. HORST, P. M. PARDALOS AND N. V. THOAI, *Introduction to Global Optimization*, Kluwer Academic Publishers, 1995 (**Second Edition** (2000)). [Chinese translation published by Tsinghua University Press, 2003.]

2 Why these problems are difficult?

■ Over an unbounded domain ($D = \mathbb{R}^n$ or $D = \mathbb{Z}^n$), no algorithm can be designed for integer programming, nonlinear equations, continuous global optimization and constraint satisfaction problems.

■ **Diophantine Equation Problem (Hilbert 23 problems):** Given a polynomial function $P(x_1, ..., x_n)$ with integer coefficients, decide whether the following equation has a solution:

$$\begin{cases} P(x_1, ..., x_n) = 0, \\ x_i \; : \; integer, \; i = 1, ..., n? \end{cases}$$

■ There exists no recursive function to decide whether the diophantine equation problem has a solution.

2 Why these problems are difficult?

The main focus of computational complexity is to analyze the **intrinsic difficulty** of optimization problems and to decide which of them are likely to be tractable. The pursuit for developing efficient algorithms also leads to **elegant general approaches** for solving optimization problems, and reveals **surprising connections** among problems and their solutions.

- The general problem is **NP-hard**. Furthermore, checking existence of a feasible point that satisfies the optimality conditions is also **NP-hard**.

- **How to check convexity!**

2 Why these problems are difficult?

Challenging problems:

- Phase transitions problems
- Average case complexity
- Smoothed Analysis

References

[1] P.M. PARDALOS, *Complexity in Numerical Optimization*, World Scientific, (1993).

[2] P.M. PARDALOS, *Approximation and Complexity in Numerical Optimization*, Kluwer Academic Publishers, (2000).

3 DC Optimization Problems

■ Many powerful techniques in global optimization are based on the fact that many objective functions can be expressed as the **difference of two convex functions** (so called **d.c. functions**).

■ If $D(x)$ is an objective function in R^n, then the representation $D(x) = p(x) - q(x)$, where p, q are convex functions is said to be a **d.c. decomposition** of D.

■ The space of d.c. functions is **closed** under many operations frequently encountered in optimization (i.e., sum, product, max, min, etc).

■ **Hartman 1959**: Every locally d.c. function is d.c.

3 DC Optimization Problems

■ For simplicity of notation, consider the d.c. program:

$$\min \quad f(x) - g(x)$$

$$\text{s.t.} \quad x \in D$$

(1)

where D is a *polytope* in R^n with nonempty interior, and f and g are *convex functions* on R^n.

■ By introducing an additional variable t, Problem (1) can be converted into the equivalent problem:

3 DC Optimization Problems

- **Global Concave Minimization**:

$$\min \quad t - g(x)$$

$$\text{s.t.} \quad x \in D, f(x) - t \leq 0 \tag{2}$$

with **concave objective function** $t - g(x)$ and **convex feasible set** $\{(x, t) \in R^{n+1} : x \in D, f(x) - t \leq 0\}$. If (x^*, t^*) is an optimal solution of (2), then x^* is an optimal solution of (1) and $t^* = f(x^*)$.

- **Therefore, any d.c. program of type (1) can be solved by an algorithm for minimizing a concave function over a convex set.**

4 DI Optimization Problems

Monotonicity with respect to some variables (partial monotonicity) or to all variables (total monotonicity) is a natural property exhibited by many problems encountered in applications. The most general problem of **d.i. monotonic optimization** is:

$$
\begin{aligned}
\min \quad & f(x) - g(x) \\
\text{s.t.} \quad & f_i(x) - g_i(x) \leq 0, i = 1, \ldots, m
\end{aligned}
\tag{3}
$$

where are all functions are increasing on R^n_+.

4 DI Optimization Problems

- Assume without loss of generality that $g(x) = 0$.
- $\{\forall i \ \ f_i(x) - g_i(x) \leq 0\} \Leftrightarrow \max_{1 \leq i \leq m} \{f_i(x) - g_i(x)\} \leq 0 \Leftrightarrow F(x) - G(x) \leq 0$, where

$$F(x) = \max_i \{f_i(x) + \sum_{i \neq j} g_j(x)\},$$

$$G(x) = \sum_i g_i(x)$$

- $F(x)$ and $G(x)$ are both increasing functions.

4 DI Optimization Problems

- Problem reduces to:

$$
\begin{aligned}
\min \quad & f(x) \\
\text{s.t.} \quad & F(x) + t \leq F(b), \\
& G(x) + t \geq F(b), \\
& 0 \leq t \leq F(b) - F(0), \\
& x \in [0, b] \subset R_+^n.
\end{aligned}
$$

- A set $G \subseteq R_+^n$ **normal** if for any two points x, x' such that $x' \leq x$, if $x \in G$, then $x' \in G$.

4 DI Optimization Problems

Numerous global optimization problems can be reformulated as monotonic optimization problems. Such problems include multiplicative programming, nonconvex quadratic programming, polynomial programming, and Lipschitz optimization problems.

References

[1] H. Tuy, *Monotonic Optimization*, SIAM Journal of Optimization Vol. 11, No. 2 (2000), pp. 464-494.

[2] P.M. Pardalos, H.E. Romeijn, H. Tuy, *Recent developments and trends in Global Optimization*, Journal of Comp.Appl. Math, 124 (2000), pp. 209–228.

5 Challenging problems

- Find the **best** D.C. decomposition.
- Find the **best** D.I. decomposition.

6 Is Continuous Optimization different than Discrete Optimization?

In combinatorial optimization and graph theory many approaches have been developed that link the discrete universe to the continuous universe through **geometric, analytic, and algebraic** techniques. Such techniques include global optimization formulations, semidefinite programming, and spectral theory.

Examples:

- **Interior Point and Semidefinite Programming Algorithms**

- **Lovász number**

- **Goemans-Williamson Relaxation of the MAX-CUT**

- **Solution of Gilbert-Pollak's Conjecture (Du-Hwang)**

6 Is Continuous Optimization different than Discrete Optimization?

Examples:

- $\boxed{z \in \{0, 1\} \Leftrightarrow z - z^2 = z(1 - z) = 0}$

- **Integer constraints are equivalent to continuous nonconvex constraints (complementarity!)**

- $\boxed{\textbf{Discrete Optimization} \Longleftrightarrow \textbf{Continuous Optimization}}$

- **The key issue is:**
  $\boxed{\textbf{Convex Optimization} \neq \textbf{Nonconvex Optimization}}$

- **The Linear complementarity problem (LCP) is equivalent to the linear mixed integer feasibility problem (Pardalos-Rosen)**

References

[1] P. M. PARDALOS AND J. B. ROSEN, *Constrained Global Optimization: Algorithms and Applications*, Springer-Verlag, Berlin, 1987.

[2] PANOS M. PARDALOS AND HENRY WOLKOWICZ, *Topics in Semidefinite and Interior-Point Methods*, American Mathematical Society (1998).

[3] P.M. PARDALOS, *Continuous Approaches to Discrete Optimization Problems*, In *Nonlinear Optimization and Applications*, G. Di Pillo & F. Giannessi, Ed., Plenum (1996), pp. 313-328.

References

[4]   J. MITCHELL, P.M. PARDALOS AND M.G.C. RESENDE, *Interior Point Methods for Combinatorial Optimization*, In *Handbook of Combinatorial Optimization* Vol. 1 (1998), pp. 189-298.

[5]   D.-Z. DU AND P.M. PARDALOS, *Global Minimax Approaches for Solving Discrete Problems*, Springer-Verlag (1997), pp. 34-48.

7.1 Satisfiability Problems

The satisfiability problem (SAT) is central in mathematical logic, computing theory, and many industrial application problems. Problems in computer vision, VLSI design, databases, automated reasoning, computer-aided design and manufacturing, involve the solution of instances of the satisfiability problem. Furthermore, SAT is the basic problem in computational complexity. Developing efficient exact algorithms and heuristics for satisfiability problems can lead to general approaches for solving combinatorial optimization problems.

7.1 Satisfiability Problems

Let $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_n$ be n clauses, involving m Boolean variables $x_1, x_2, \ldots, x_m$, which can take on only the values `true` or `false` (1 or 0). Define clause i to be

$$\mathcal{C}_i = \bigvee_{j=1}^{m_i} l_{ij},$$

where the literals $l_{ij} \in \{x_i, \bar{x}_i \mid i = 1, \ldots, m\}$.

In the **Satisfiability Problem** (CNF)

$$\bigwedge_{i=1}^{n} \mathcal{C}_i = \bigwedge_{i=1}^{n} (\bigvee_{j=1}^{m_i} l_{ij})$$

one is to determine the assignment of truth values to the m variables that satisfy all n clauses.

7.1 Satisfiability Problems

Given a CNF formula $F(\mathbf{x})$ from $\{0,1\}^m$ to $\{0,1\}$ with n clauses $C_1, \ldots, C_n$, we define a real function $f(\mathbf{y})$ from E^m to E that transforms the SAT problem into an unconstrained **global optimization problem**:

$$\min_{\mathbf{y} \in \mathbf{E}^m} f(\mathbf{y}) \tag{4}$$

where

$$f(\mathbf{y}) = \sum_{i=1}^{n} c_i(\mathbf{y}). \tag{5}$$

A clause function $c_i(\mathbf{y})$ is a product of m *literal functions* $q_{ij}(y_j)$ $(1 \leq j \leq m)$:

$$c_i = \prod_{j=1}^{m} q_{ij}(y_j), \tag{6}$$

7.1 Satisfiability Problems

where

$$q_{ij}(y_j) = \begin{cases} |y_j - 1| & \text{if literal } x_j \text{ is in clause } C_i \\ |y_j + 1| & \text{if literal } \bar{x}_j \text{ is in clause } C_i \\ 1 & \text{if neither } x_j \text{ nor } \bar{x}_j \text{ is in } C_i \end{cases} \tag{7}$$

The correspondence between $\mathbf{x}$ and $\mathbf{y}$ is defined as follows (for $1 \le i \le m$):

$$x_i = \begin{cases} 1 & \text{if } y_i = 1 \\ 0 & \text{if } y_i = -1 \\ undefined & \text{otherwise} \end{cases}$$

$F(\mathbf{x})$ is true iff $f(\mathbf{y}){=}0$ on the corresponding $\mathbf{y} \in \{-1, 1\}^m$.

7.1 Satisfiability Problems

Next consider a polynomial unconstrained **global optimization** formulation:

$$\min_{\mathbf{y} \in \mathbf{E}^m} f(\mathbf{y}), \tag{8}$$

where

$$f(\mathbf{y}) = \sum_{i=1}^{n} c_i(\mathbf{y}). \tag{9}$$

A clause function $c_i(\mathbf{y})$ is a product of m *literal functions* $q_{ij}(y_j)$ $(1 \leq j \leq m)$:

$$c_i = \prod_{j=1}^{m} q_{ij}(y_j), \tag{10}$$

7.1 Satisfiability Problems

where

$$q_{ij}(y_j) = \begin{cases} (y_j - 1)^{2p} & \text{if } x_j \text{ is in clause } C_i \\ (y_j + 1)^{2p} & \text{if } \bar{x}_j \text{ is in clause } C_i \\ 1 & \text{if neither } x_j \text{ nor } \bar{x}_j \text{ is in } C_i \end{cases} \tag{11}$$

where p is a positive integer.

The correspondence between $\mathbf{x}$ and $\mathbf{y}$ is defined as follows (for $1 \leq i \leq m$):

$$x_i = \begin{cases} 1 & \text{if } y_i = 1 \\ 0 & \text{if } y_i = -1 \\ undefined & \text{otherwise} \end{cases}$$

$F(\mathbf{x})$ is true iff $f(\mathbf{y})=0$ on the corresponding $\mathbf{y} \in \{-1, 1\}^m$.

7.1 Satisfiability Problems

- **These models transform the SAT problem from a discrete, constrained decision problem into an unconstrained global optimization problem.**

- A good property of the transformation is that **these models establish a correspondence between the global minimum points of the objective function and the solutions of the original SAT problem.**

- A *CNF* $F(\mathbf{x})$ is true *if and only if* f takes the global minimum value 0 on the corresponding $\mathbf{y}$.

References

[1] D.-Z. DU, J. GU, AND P. M. PARDALOS (Editors),
     Satisfiability Problem: Theory and Applications, *DIMACS
     Series Vol. 35, American Mathematical Society, (1997).*

7.2 The Maximum Clique Problem

Consider a graph $G = G(V, E)$, where $V = \{1, \ldots, n\}$ denotes the set of vertices (nodes), and E denotes the set of edges. Denote by (i, j) an edge joining vertex i and vertex j. A clique of G is a subset C of vertices with the property that every pair of vertices in C is joined by an edge. The maximum clique problem is the problem of finding a clique set C of maximal cardinality.

Applications:
● project selection, classification theory, fault tolerance, coding theory, computer vision, economics, information retrieval, signal transmission theory, aligning DNA and protein sequences, and other specific problems.

Multivariable polynomial formulations

- If x^* is the solution of the following (**continuous**) quadratic program

$$\max f(x) = \sum_{i=1}^{n} x_i - \sum_{(i,j)\in E} x_i x_j = e^T x - 1/2 x^T A_G x$$
$$\text{subject to } 0 \le x_i \le 1 \text{ for all } 1 \le i \le n$$

  then, $f(x^*)$ equals the size of the maximum independent set.

- If x^* is the solution of the following (**continuous**) polynomial program

$$\max f(x) = \sum_{i=1}^{n} (1 - x_i) \prod_{(i,j)\in E} x_j$$
$$\text{subject to } 0 \le x_i \le 1 \text{ for all } 1 \le i \le n$$

  then, $f(x^*)$ equals the size of the maximum independent set.

- **In both cases a polynomial time algorithm has been developed that finds independent sets of large size.**

Multivariable polynomial formulations

References

[1] J. Abello, S. Butenko, P. M. Pardalos and M. G. C. Resende, "Finding Independent Sets in a Graph Using Continuous Multivariable Polynomial Formulations", *Journal of Global Optimization 21* (2001), pp. 111-137.

Motzkin-Strauss type approaches

Consider the continuous **indefinite quadratic programming problem**

$$\max \quad f_G(x) = \sum_{(i,j) \in E} x_i x_j = \tfrac{1}{2} x^T A_G x$$

$$\text{s.t.} \quad x \in S = \{x = (x_1, \ldots, x_n)^T : \sum_{i=1}^{n} x_i = 1, \quad (12)$$

$$x_i \geq 0 \quad (i = 1, \ldots, n)\},$$

where A_G is the adjacency matrix of the graph G.

Motzkin-Strauss type approaches

- If $\alpha = \max\{f_G(x) : x \in S\}$, then G has a maximum clique C of size $\omega(G) = 1/(1 - 2\alpha)$. This maximum can be attained by setting $x_i = 1/k$ if $i \in C$ and $x_i = 0$ if $i \notin C$.

- (Pardalos and Phillips 1990) If A_G has r negative eigenvalues, then at least $n - r$ constraints are active at any global maximum x^* of $f(x)$. Therefore, if A_G has r negative eigenvalues, then the size $|C|$ of the maximum clique is bounded by $|C| \leq r + 1$.

The Call Graph

The **"call graph"** comes from telecommunications traffic. The vertices of this graph are telephone numbers, and the edges are calls made from one number to another (including additional billing data, such as, the time of the call and its duration). The challenge in studying call graphs is that they are massive. Every day AT & T handles approximately 300 million long-distance calls.

Careful analysis of the call graph could help with infrastructure planning, customer classification and marketing.

How can we visualize such massive graphs? To flash a terabyte of data on a 1000x1000 screen, you need to cram a megabyte of data into each pixel!

Recent Work on Massive Telecommunication Graphs

In our experiments with data from **telecommunication traffic**, the corresponding multigraph has **53,767,087 vertices and over 170 million of edges**.

A **giant connected component with 44,989,297** vertices was computed. The **maximum (quasi)-clique problem** is considered in this giant component.

References

[1] J. Abello, P.M. Pardalos and M.G.C. Resende, "*On maximum clique problems in very large graphs*", In **DIMACS** Vol. 50 (1999), American Mathematical Society, pp. 119-130.

[2] American Scientist, Jan./ Febr. (2000) and Sept./Oct. (2006).

Optimization on Massive Graphs

Several other graphs have been considered:

- Financial graphs
- Brain models
- Drug design models
- Biological networks

Optimization on Massive Graphs

J. Abello, P. M. Pardalos and M.G.C. Resende (Editors), "Handbook of Massive Data Sets ", Kluwer Academic Publishers, Dordrecht, 2002.

V. Boginski, S. Butenko, and P.M. Pardalos, On Structural Properties of the Market Graph, In "Innovations in Financial and Economic Networks" (A. Nagurney, Ed.), Edward Elgar Publishers, 2003, pp. 29-45.

7.3 Multiquadratic Optimization
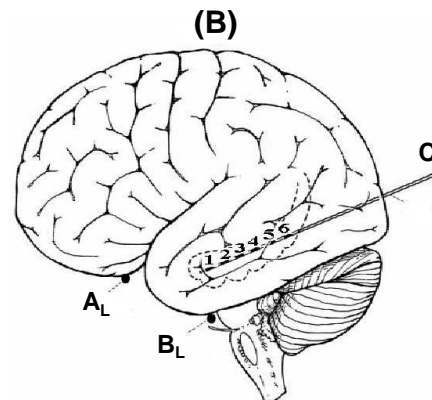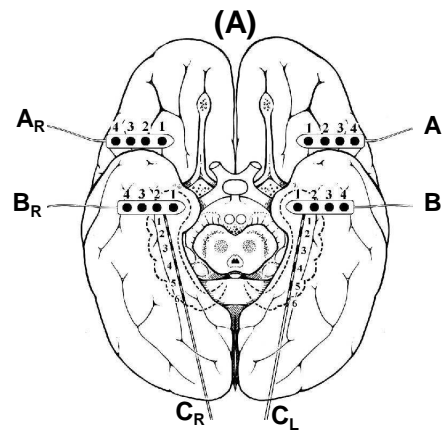
7.3 Multiquadratic Optimization

■ **Estimation of Short Term Largest Lyapunov Exponents (STL_{max})**

Since **the brain is a nonstationary system**, algorithms used to estimate measures of the brain dynamics should be capable of automatically identifying and appropriately weighing existing transients in the data. In a chaotic system, orbits originating from similar initial conditions (nearby points in the state space) diverge exponentially (expansion process). The **rate of divergence** is an important aspect of the system dynamics and is reflected in the value of **Lyapunov exponents**.

7.3 Multiquadratic Optimization

- **Spatiotemporal Dynamical Analysis**

We employ the T-index (from the well-known paired T-statistics for comparisons of means) as a measure of distance between the mean values of pairs of STL_{max} profiles over time. The T-index at time t between electrode sites i and j is defined as:

$$T_{i,j}(t) = \sqrt{N} \times |E\{STL_{max,i} - STL_{max,j}\}|/\sigma_{i,j}(t)$$

where $E\{\cdot\}$ is the sample average difference for the $STL_{max,i} - STL_{max,j}$ estimated over a moving window $w_t(\lambda)$.
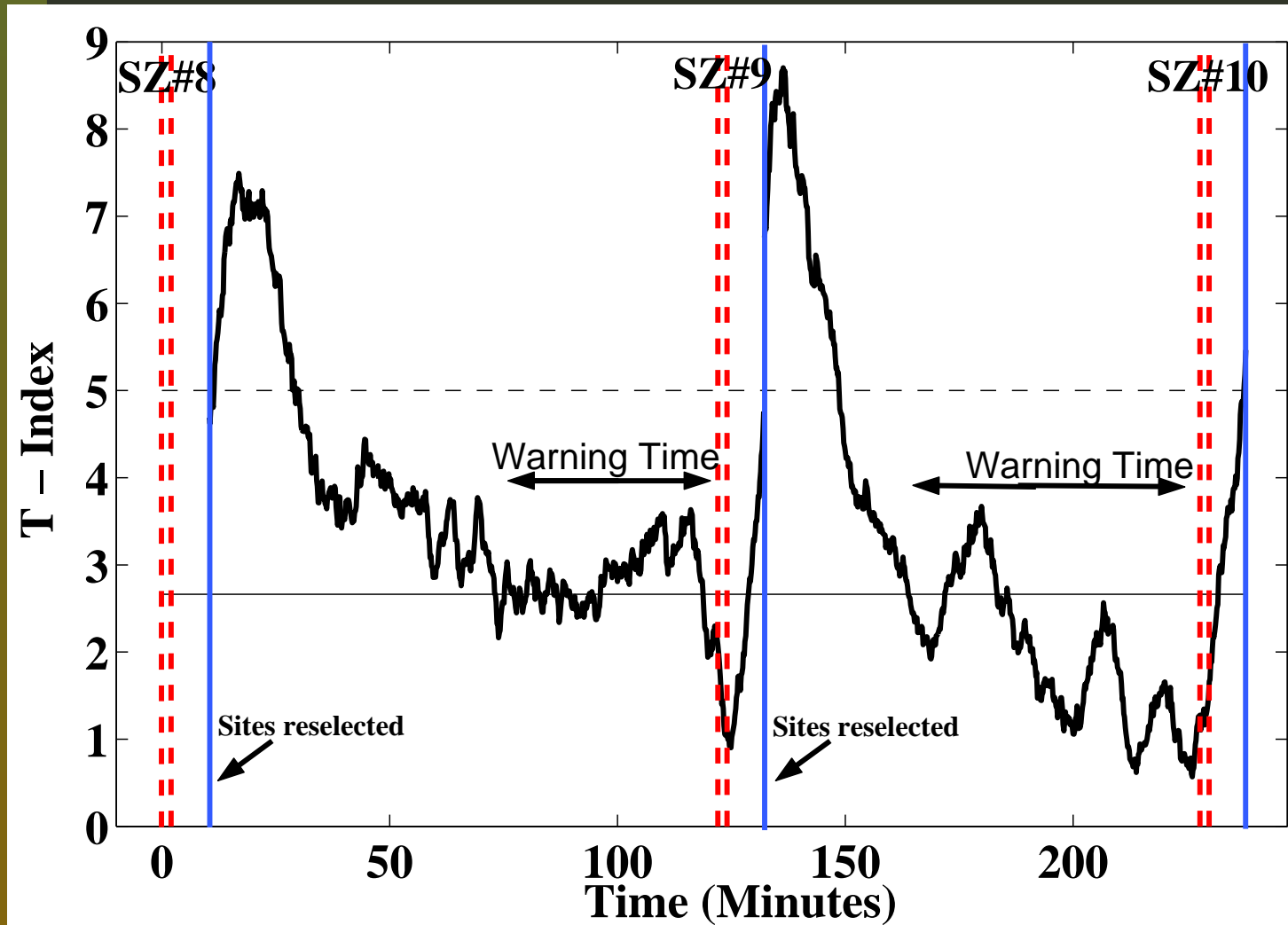
7.3 Multiquadratic Optimization

■ **Electrode Selection Problem**

The quadratically constrained quadratic 0-1 problem is given by:

$$
\begin{aligned}
\min \quad & x^T A x \\
\text{s.t.} \quad & \sum_{i=1}^{n} x_i = k \\
& x^T B x \geq T_\alpha k(k-1) \\
& x \in \{0,1\}^n
\end{aligned}
$$

7.3 Multiquadratic Optimization

7.3 Multiquadratic Optimization

References

[1] P.M. Pardalos, L.D. Iasemidis, D.-S. Shiau, J.C. Sackellares, **Quadratic binary programming and dynamic system approach to determine the predictability of epileptic seizures** (with L.D. Iasemidis, D.-S. Shiau, J.C. Sackellares), Journal of Combinatorial Optimization Vol. 5 No. 1 (2001), pp. 9-26.

[2] P.M. Pardalos, L.D. Iasemidis, J.C. Sackellares, D.-S. Shiau, W. Chaovalitwongse, P.R. Carney, J.C. Principe, M.C.K. Yang, V.A. Yatsenko , S.N. Roper, **Seizure Warning Algorithm Based on Optimization and Nonlinear Dynamics**, Mathematical Programming, Vol. 101/2 (2004), pp. 365–385. **The William Pierskalla best paper award for research excellence in health care management science, INFORMS.**

Biocomputing

- Biocomputing, co-editors: P.M. Pardalos and J. Principe, Kluwer Academic Publishers, (2002).

- Quantitative Neuroscience, co-editors: P.M. Pardalos, C. Sackellares, P. Carney, and L. Iasemidis, Kluwer Academic Publishers, (2004).

- Data Mining in Biomedicine, co-editors: P.M. Pardalos, V. Boginski, and A. Vazacopoulos, Springer, (2007).

8 Hierarchical Optimization

8 Hierarchical Optimization

- **The word *hierarchy* comes from the Greek word "$\iota\epsilon\rho\alpha\rho\chi\iota\alpha$", a system of graded (religious) authority**.

8 Hierarchical Optimization

- **The word *hierarchy* comes from the Greek word "$\iota\epsilon\rho\alpha\rho\chi\iota\alpha$", a system of graded (religious) authority**.

- **Hierarchical structures** are found in many complex systems and in particular in biology.

- **Biological systems** are characterized by hierarchical architectural designs in which organization is controlled on length scales ranging from the molecular to macroscopic. These hierarchical architectures rely on critical interfaces that link structural elements of disparate scale.

8 Hierarchical Optimization

Nature makes very different systems (that have specific hierarchical composite structures) out of very similar molecular constituents.

8 Hierarchical Optimization

Nature makes very different systems (that have specific hierarchical composite structures) out of very similar molecular constituents.

■ First, the structures are organized in discrete levels.

8 Hierarchical Optimization

Nature makes very different systems (that have specific hierarchical composite structures) out of very similar molecular constituents.

- First, the structures are organized in discrete levels.

- Second, the levels of structural organization are held together by specific interactions between components.

8 Hierarchical Optimization

Nature makes very different systems (that have specific hierarchical composite structures) out of very similar molecular constituents.

- First, the structures are organized in discrete levels.
- Second, the levels of structural organization are held together by specific interactions between components.
- Finally, these interacting levels are organized into an oriented distinct hierarchical composite system of specific function.

8 Hierarchical Optimization

The mathematical study of hierarchical structures can be found in diverse scientific disciplines including environment, ecology, biology, chemical engineering, classification theory, databases, network design, game theory and economics. The study of hierarchy occurring in biological structures reveals interesting properties as well as limitations due to different properties of molecules. Understanding the complexity of hierarchical designs requires "**systems methodologies that are amenable to modeling, analyzing and optimizing**" (Haimes Y.Y. 1977) these structures.

8 Hierarchical Optimization

- **Hierarchical optimization (or multi-level)** can be used to study properties of these hierarchical designs. In **hierarchical optimization, the constraint domain is implicitly determined by a series of optimization problems which must be solved in a predetermined sequence**.

- Hierarchical optimization is a generalization of mathematical programming. The simplest two-level (or bilevel) programming problem describes a hierarchical system which is composed of two levels of decision makers and is stated as follows:

8 Hierarchical Optimization

$$\textbf{(BP)} \min_{y \in Y} \quad \varphi(x(y), y) \tag{13}$$

$$\text{subject to} \quad \psi(x(y), y) \leq 0 \tag{14}$$

$$\text{where} \quad x(y) \; = \; \arg\min_{x \in X} f(x, y) \tag{15}$$

$$\text{subject to} \quad g(x, y) \leq 0, \tag{16}$$

where $X \subset R^n$ and $Y \subset R^m$ are closed sets, $\psi : X \times Y \rightarrow R^p$ and $g : X \times Y \rightarrow R^q$ are multifunctions, φ and f are real-valued functions. The set
$\mathcal{S} = \{(x, y) : x \in X, y \in Y, \psi(x, y) \leq 0, g(x, y) \leq 0\}$ is the *constraint set* of **BP**.

8 Hierarchical Optimization

■ Multi-level programming problems have been studied extensively in their general setting during the last decade.

■ In general, hierarchical optimization problems are nonconvex and therefore is not easy to find globally optimal solutions.

8 Hierarchical Optimization

- It seems that **hierarchical structures are harder to manage than completely centralized systems**. Then, **what are the rationalities for hierarchical structures to exist**?

- Answers to such questions may help us to understand the reason behind hierarchical structures in biology.

8 Hierarchical Optimization

References

[1]   A. Migdalas, P.M. Pardalos, and P. Varbrand (Editors), **Multilevel optimization: Algorithms and Applications**, NOIA vol 20, Kluwer Academic Publishers, 1997.

[2]   S. Dempe, **Foundations of Bilevel Programming**, NOIA vol 61, Kluwer Academic Publishers, 2002.

Algorithms

- extreme point algorithms

- branch-and-bound algorithms

- complementarity pivot algorithms

- descent methods

- penalty function methods

Surveys

References

[3]   L.N. Vicente, P.H. Calamai, *Bilevel and Multilevel Programming: A Bibliography Review*, Journal of Global Optimization **5**: 291–306, 1994.

[4]   S. Dempe, *Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints*, Optimization **52**, 333–359, 2003.

[5]   B. Colson, P. Marcotte, G. Savard, *Bilevel Programming: A Survey*, 4OR **3**: 87–107, 2005.

Recent work

- A. Chinchuluun, P.M. Pardalos, and H.X. Huang. Multilevel Optimization: Complexity Issues, Optimality Conditions, Algorithms. In: Gao, D.Y. and Sherali, H.D., editors, Advances in Mechanics and Mathematics, Vol. 3, pages 197-221 (2008).

- A. Chinchuluun, P.M. Pardalos, A. Migdalas, L. Pitsoulis, editors. Pareto Optimality, Game Theory and Equilibria, Springer, 2008

Multivariate Partition Approach

Multivariate Partition Approach

- This is an approach for solving a single-level mathematical programming problem based on its equivalent multilevel programming formulation.

- The basic idea of this approach is to partition all the variables appearing in the optimization problem into two or more groups and then, solving some small-sized problems with respect to each group of variables, to generate an *improved exploratory approximate solution* of the initial problem.

- Based on the improved exploratory approximate solution, new partition groups of variables can be constructed, and the corresponding small-sized problems can be solved as before in order to obtain a better approximate solution.

Multivariate Partition Approach

- **With this approach we can formulate optimization problems as multi-level optimization problems.**

- In fact, there are well-known algorithms which are related or can be regarded as specific implementations of the MPA.

Multivariate Partition Approach

- Consider the following problem:

$$\min_{x \in D \subseteq R^n} f(x), \qquad (1)$$

where D is a robust set and $f(x)$ is continuous in D.

- Let $\{\Delta_i, i = 1, \ldots, p\}$ be a partition of $S = \{x_1, \ldots, x_n\}, p > 1$.

Multivariate Partition Approach

- (1) is equivalent to the following multilevel optimization problem:

$$\min_{y_{\sigma_1} \in D_{\sigma_1}} \left\{ \min_{y_{\sigma_2} \in D_{\sigma_2}} \ldots \left\{ \min_{y_{\sigma_p} \in D_{\sigma_p}} f(\Delta_1, \ldots, \Delta_p) \right\} \ldots \right\},$$

$$(2)$$

where $\sigma = (\sigma_1, \ldots, \sigma_n)$ is any permutation of $\{1, 2, \ldots, p\}$. The components of the vector y_{σ_i} coincide with the elements of Δ_i and D_{σ_i} is defined as a feasible domain of y_{σ_i}.

Multivariate Partition Approach

Example. Let $D = \{x : \ \|x\| \leq 1\} \subset \mathbb{R}^2$.

- Let the partition of $S = \{x_1, x_2\}$ be $\{\Delta_1, \Delta_2\}$, where $\Delta_1 = \{x_1\}$ and $\Delta_2 = \{x_2\}$.

- If the permutation $\sigma = (2, 1)$ then $D_\sigma = (D_2, D_1)$, where $D_2 = [-1, 1]$, $D_1 = \left[-\sqrt{1 - x_2^2}, \sqrt{1 - x_2^2} \right]$

- Aggregated variables are $y_2 = x_2 \in D_2$ and $y_1 = x_1 \in D_1$.

Multivariate Partition Approach

- Among various multilevel formulations of problem (1), the bilevel optimization formulation seems to be most useful to the implementation of the MPA.

- Let $S = \{x_1, \ldots, x_n\}$, and let $\{I^-, I^+\}$ be a partition of an index set $I_n = \{1, \ldots, n\}$.

- Denote by S^- and S^+ the corresponding subset of S with respect to I^- and I^+, respectively.

- $\{S^-, S^+\}$ is a partition of the set S.

Multivariate Partition Approach

The bilevel optimization formulation is as follows:

$$\min_{y^- \in D_{S^-}} g(S^-),$$

where $g(S^-)$ is the global minimum of the problem

$$\min_{y^+ \in D_{S^+}} f(S^-, S^+)$$

- I^- and I^+ are called the *reference index set* and the *active index set*. S^- and S^+ are called the set of *reference variables* and the set of *active variables*.

Multivariate Partition Approach

- For any partition $\{\Delta_i | i = 1, \ldots, p\}$ of the set S and any permutation $\sigma = (\sigma_1, \ldots, \sigma_n)$ of the index set I_p, let $y_\sigma = (y_{\sigma_1}, \ldots, y_{\sigma_p})$ be a feasible point with respect to the domain vector $D_\sigma = (D_{\sigma_1}, \ldots, D_{\sigma_p})$.

- By the *the multivariate partition approach (MPA)* to the initial problem (1) we mean the method that improves the performance of the feasible point y_σ using the following strategy:

Multivariate Partition Approach

- Let $S^- = S \backslash \Delta_{\sigma_i}$ be a set of reference variables for each $i \in I_p$. Solving the followers' problem in the above bilevel formulation we can obtain an improved value $\tilde{y}_{\sigma_i}$ of the active variable

$$y^+ = y_{\sigma_i} \in D_{S^+} = D_{\sigma_i}^+ \subseteq D_{\sigma_i}$$

- **Improved Point.** Based on the improved point $\tilde{y}_p = (\tilde{y}_{\sigma_1}, \dots, \tilde{y}_{\sigma_p})$ and the feasible point y_σ, an improved feasible point y_σ^* can be determined by a search and decision scheme.

Multivariate Partition Approach

- Under some conditions it can be proved that the MPA converges.

- H. X. Huang and P.M. Pardalos, **Multivariate Partition Approach for Optimization Problems**, Cybernetics and Systems Analysis Vol. 38, No. 2 (2002), pp. 265-275.

Examples of the MPA

■ K-Means Type Algorithms

■ Coordinate Descent Method

K-Means Algorithm

- The K-means method (MacQueen, 1967) is one of the most popular clustering methods which have been applied in a variety of fields including pattern recognition, information retrieval, document extraction, and microbiology analysis, etc.

- The goal of this method is to classify a given data set through a certain number of clusters, such that some metric relative to the centroids of the clusters is minimized.

K-Means Algorithm

We can define our problem mathematically as follows:

- We are given a set X of a finite number of points in $\mathbb{R}^n$:

$$X = (x^1, x^2, \ldots, x^n) \text{ where } x^i \in \mathbb{R}^d, \ i = 1, 2, \ldots, n.$$

- We aim at finding a partition $C_j \neq \emptyset$, $j = 1, 2, \ldots, k$:

$$X = \bigcup_{j=1}^{k} C_j, \ C_j \cap C_l = \emptyset \text{ for all } j \neq l,$$

of X, which minimizes the squared error function:

K-Means Algorithm

$$f(C_1, C_2, \ldots, C_k) = g(c^1, c^2, \ldots, c^k) = \sum_{j=1}^{k} \sum_{x^i \in C_j} \|c^j - x^i\|^2,$$

where c^j is the center of the cluster C_j defined by

$$c^j = \frac{1}{|C_j|} \sum_{x^i \in C_j} x^i, \; j = 1, 2, \ldots, k. \tag{17}$$

- We note that, in optimal solutions, a cluster can be represented by its centroid since each point should be assigned to the cluster that has the closest centroid. Thus, we can think that variables of our optimization problem are centroids of the clusters.

K-Means Algorithm

Algorithm K-Means

Step 1. Initialize the centroids c_0^j, $j = 1, 2, \ldots, k$. Set $q = 0$.

Step 2. Assign each point x^i ($i = 1, 2, \ldots, n$) to the cluster that has the closest centroid c^j ($j \in \{1, 2, \ldots, k\}$), that is

$$j = \operatorname{argmin}_{1 \leq l \leq k} \|x^i - c_q^l\|^2.$$

Step 3. When all points have been assigned, for $j = 1, 2, \ldots, k$, calculate the new position c_{q+1}^j of the centroid j using Equation (17).

Step 4. If $c_q^j = c_{q+1}^j$ for all $j = 1, 2, \ldots, k$, then stop, otherwise set $q = q + 1$ and go to Step 2.

K-Means Algorithm

- We note that Equation (17) finds a local solution of problem

$$\min \quad g(c^1, \ldots, c^k)$$
$$\text{s.t.} \quad c^j \in \mathbb{R}^d$$

  while all other centroids (except centroid j) are fixed. Therefore, we can see that the K-means algorithm is a special case of the multivariate partition approach.

- One way to find a solution, which is close to a global minima,for the above problem is to locate centroid j at $x^1, \ldots, x^n$ and choose the one with the least value of g among them after reassigning the points to the centroids. This idea is closely related to j-means clustering algorithm (Hansen and Mladenović, 2001).

K-Means Algorithm

References

[1] J.B. MacQueen, *Some Methods for Classification and Analysis of Multivariate Observations*, In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, **1**: 281–297 (1967)

[2] P. Hansen, and N, Mladenović, *J-Means: A New Local Search Heuristic for Minimum Sum of Clustering*, Pattern Recognition, **34**, 405-413 (2001)

Coordinate Descent Method

■ If $n = p$ or $\Delta_i = \{x_i\}$ for $i = 1, \ldots, n$ in the partition of the multivariate partition approach, then many existing optimization algorithms designed for solving Problem (1) are closely related to this partition including the *coordinate descent method*.

■ Basically, each coordinate axis is searched, and a descent is only made along a unit vector. The cyclic Coordinate Descent Method minimizes a function $f(x_1, \ldots, x_n)$ cyclically with respect to the coordinate variables.

■ That is, first x_1 is searched, then x_2, etc. Various variations are possible. One advantage of these methods is their easy implementation.

Coordinate Descent Method

Meneses et al. (2005) proposed a GRASP (Greedy Randomized Search Procedure), which is a special case of coordinate descent method, for solving the following nonlinear programming problem with box constraint:

$$\min \quad f(x)$$
$$\text{s.t.} \quad l \leq x \leq u,$$

where $f(x) : \mathbb{R}^n \longrightarrow \mathbb{R}$, and l, u are lower and upper bounds for the values of the variable x.

Coordinate Descent Method

- GRASP starts with an initial point $x = \frac{l+u}{2}$.

- Then, for each $i = 1, \ldots, n$, we solve a problem of one variable x_i while other coordinate values of the current solution x are fixed. These problems are solved by discretizing the solution space using the grid density h.

- Based on solutions to these problems, the method finds an improved feasible point. Then the above procedure will be repeated.

- If the number of iterations with no improvements achieves value N (some positive integer number), the value of h is reduced to its half, and the process is restarted.

Coordinate Descent Method

References

[1] H.X. HUANG, P.M. PARDALOS, AND Z.J. SHEN, *A point balance algorithm for the spherical code problem*, Journal of Global Optimization Vol. 19, No. 4 (2001), pp. 329–344.

[2] C.N. MENESES, P.M. PARDALOS, AND M.G.C. RESENDE, *GRASP for Nonlinear Optimization*, Optimization Letters, submitted (2005)

Applications of the MPA

- Lennard-Jones Problem

- Spherical Code Problem

Lennard-Jones Problem

Problem: Compute the native 3 dimensional conformation (folded state) of a (globular) protein given its amino acid sequence, possibly in the presence of additional agents (e.g., drugs).

Lennard-Jones Problem

Given a cluster of N atoms in 3–dimensional space, the potential energy function of the cluster is defined as the summation (over all of pairs) of the two-body interatomic pair potentials. Let the code $P_N = \{x_1, \cdots, x_N\}$ be the collection of centers of N atoms. The potential energy function is defined as follows:

$$V(P_N) = \sum_{1 \leq i < j \leq N} v(\|x_i - x_j\|), \qquad (18)$$

where $\|.\|$ is the Euclidean norm and $v(r)$ is the interatomic pair potential.

Lennard-Jones Problem

- There are many potential functions that have been considered, for example, the **Lennard-Jones** potential energy function:

$$v(r) = r^{-12} - 2r^{-6}$$

- The global minimization of potential energy functions plays an important role in the determination of ground states or stable states of certain classes of molecular clusters and proteins.

Lennard-Jones Problem

- Let $P_N = \{x_1, \cdots, x_N\}$ be a code in which points are mutually different.

- We partition index set I into two disjoint sets I^+ and I^-.

- It is easy to observe that

$$V(P_N) = V(P_N(I^+)) +$$

$$+V(P_N(I^+), P_N(I^-)) + V(P_N(I^-))$$

Lennard-Jones Problem

- where

$$V(P_N(I^\pm)) = \sum_{i,j \in I^\pm, i<j} \|x_i - x_j\|^{-12} -$$

$$-2 \sum_{i,j \in I^\pm, i<j} \|x_i - x_j\|^{-6}$$

- and

$$V(P_N(I^+), P_N(I^-)) = \sum_{i \in I^+, j \in I^-} V(\{x_i, x_j\})$$

Lennard-Jones Problem

- We can consider an equivalent bilevel optimization problem

$$\min\{V(P_N(I^-)) + \bar{V}(P_N(I^-))\},$$

$$\text{s.t. } x_i \in \mathbb{R}^3, \ \forall i \in I^-$$

where $\bar{V}(P_N(I^-))\}$ is the globally optimal solution of the sub-problem:

$$\min\{V(P_N(I^+)) + V(P_N(I^+), P_N(I^-))\}$$

$$\text{s.t. } x_j \in \mathbb{R}^3, \ \forall j \in I^+$$

Lennard-Jones Problem

References

[1] H.X. HUANG, P.M. PARDALOS, AND Z.J. SHEN, *Equivalent formulations and necessary optimality conditions for the Lenard-Jones problem*, Journal of Global Optimization Vol. 22, (2002), pp. 97–118.

Spherical Code Problem

- **The Spherical Code (SC) Problem** is referred to as how to distribute points on the unit sphere according to a certain 'generalized energy'.

- The SC problem has been the focus of research in various fields such as physics, molecular biology, signal transmission, chemistry.

- Some well-known problems such as *the Tammes problem*, *the Fekete problem* and *the maximum volume arrangements* can be considered as SC problems with different objectives.

Spherical Code Problem

The Tammes problem is defined as how to place N points on a sphere in the n-dimensional space so as to maximize the minimum distance (or equivalently the minimum angle) between any two points.

$$E_T = \max_{||x_i||=1,\ 1 \leq i \leq N} \ \min_{1 \leq j < k \leq N} ||x_j - x_k||$$

Spherical Code Problem

- Let $S^n = \{x \mid \|x\| = 1\}$ denote the unit sphere in $\mathbb{R}^n$.

- Let P_N denote a spherical code with N points on S^N, i.e. $P_N = \{x_1, \ldots, x_N\}$

- *S-energy* associated with the spherical code P_N is defined as

$$
w(s, P_N) = \begin{cases} \sum_{i<j} \|x_i - x_j\|^{-s} & \text{if } s \neq 0 \\ \sum_{i<j} \ln\left(\frac{1}{\|x_i - x_j\|}\right) & \text{if } s = 0. \end{cases}
$$

Spherical Code Problem

- The *s-extremal energy* $E_N(s)$ for N points on S^n is defined by

$$E_N(s) = \begin{cases} \min_{P_N \subset S^n} w(s, P_N) \text{ if } s \geq 0 \\ \max_{P_N \subset S^n} w(s, P_N) \text{ if } s < 0 \end{cases}$$

- The spherical code $P_{s,N}^*$ is called the *s-extremal spherical code* if it satisfies

$$w(s, P_{s,N}^*) = E_N(s)$$

Spherical Code Problem

- The points in the 0-extremal spherical code $P^*_{0,N}$ are called *elliptic Fekete points*. Clearly, these points maximize the product of the distances between any two points in P_N, i.e.

$$\max_{\|x_i\|=1,\ 1\leq i\leq N}\ \Pi_{\,1\leq j<k\leq N}\|x_j - x_k\|$$

- The points in the 1-extremal spherical code $P^*_{1,N}$ are called *Fekete points*. The *Fekete points* in the $\mathbb{R}^3$ represent the locations of N charged particles on the unit sphere that repel each other according to Coulomb's law.

Spherical Code Problem

- A *s-extremal spherical code* is a globally optimal solution of the following problem:

$$\min f_s(P_N) \triangleq f_s(x_1, \ldots, x_N)$$

$$\text{s.t. } x_i \in S^n, \ i = 1, \ldots, n.$$

where

$$f_s(P_N) = \begin{cases} w(s, P_N) \text{ if } s \geq 0 \\ -w(s, P_N) \text{ if } s < 0 \end{cases}$$

Spherical Code Problem

- We can consider an equivalent bilevel optimization problem

$$\min\{f_s(P_N(I^-)) + \bar{f}_s(P_N(I^-))\},$$

$$\text{s.t. } x_i \in S^n, \ \forall i \in I^-$$

where $\bar{f}_s(P_N(I^-))\}$ is the globally optimal solution of the sub-problem:

$$\min\{f_s(P_N(I^+)) + f_s(P_N(I^+), P_N(I^-))\}$$

$$\text{s.t. } x_j \in S^n, \ \forall j \in I^+$$

Spherical Code Problem

References

[1] H.X. HUANG, P.M. PARDALOS, AND Z.J. SHEN,
*A point balance algorithm for the spherical code
problem*, Journal of Global Optimization Vol. 19,
No. 4 (2001), pp. 329–344.

HERACLITUS

"Seekers after gold dig up much earth and find little"

"The lord whose oracle is at Delphi neither speaks nor conceals, but gives signs"

- HERACLITUS