A Multigrid Optimization Framework for Centroidal Voronoi Tessellation

Zichao Di
Department of Mathematical Sciences
George Mason University

Collaborators:
Dr. Stephen G. Nash
Department of Systems Engineering & Operations Research, GMU
Dr. Maria Emelianenko
Department of Mathematical Sciences, GMU

Outline

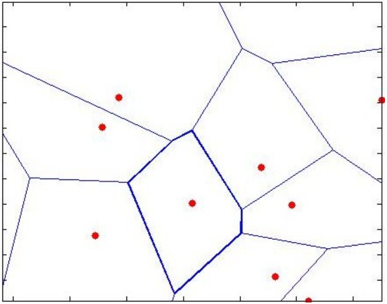
- CVT: introduction
  - CVT: concepts
  - List of applications
  - Some properties of CVTs
- Lloyd acceleration techniques
  - Lloyd method
  - Convergence result
  - Acceleration schemes
- An overview of multigrid optimization (MG/OPT) framework
  - Algorithm Description
  - Convergence and Descent
- Applying MG/OPT to the CVT Formulation
- Numerical Experiments
  - 1-dimensional examples
  - 2-dimensional examples
- Comments and Conclusions

Concept of the Voronoi tessellation

- *Given*
  - a set S
  - *elements* $z_i, i = 1, 2, \ldots, K$
  - *a distance function* $d(z, w), \forall z, w \in S$
- *The Voronoi set* V_j *is the set of all elements belonging to* S *that are closer to* z_j *than to any of the other elements* z_i*, that is*

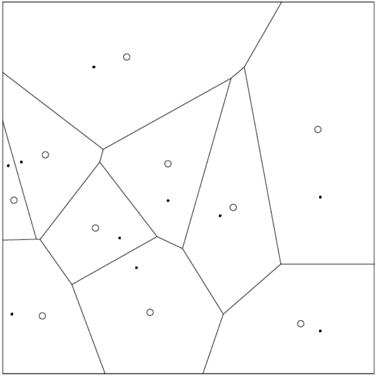
$$V_j = \{w \in S \mid d(w, z_j) < d(w, z_i),\ i = 1, \ldots, K, i \neq j\}$$

- $\{V_1, V_2, \ldots, V_k\}$ *is a Voronoi tessellation of* S
- $\{z_i\}$ *are generators of the Voronoi tessellation*



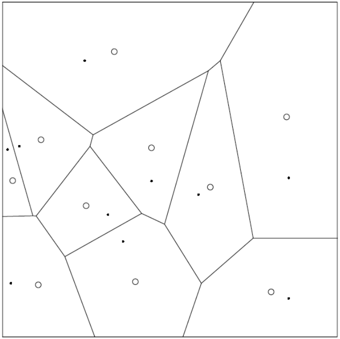
CVT: facts and definitions

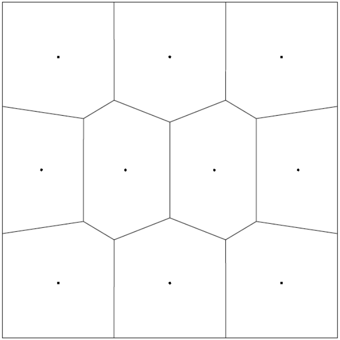
- Given the Voronoi tessellation $\{V_i\}$ corresponding to the generators $\{z_i\}$
  - The associated centroids $z_i^* = \dfrac{\int_{V_i} \rho(y) y dy}{\int_{V_i} \rho(y) dy}$, $i = 1, \dots, K$



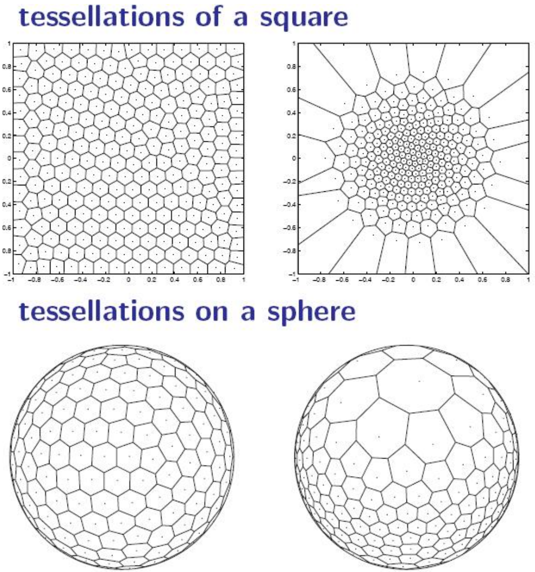
CVT: facts and definitions

- Given the Voronoi tessellation $\{V_i\}$ corresponding to the generators $\{z_i\}$, the associated centroids $z_i^* = \dfrac{\int_{V_i} \rho(y) y dy}{\int_{V_i} \rho(y) dy}$, $i = 1, \ldots, K$
- If $z_i = z_i^*$, $i = 1, \ldots, K$ we call this kind of tessellation Centroidal Voronoi Tessellation (CVT)





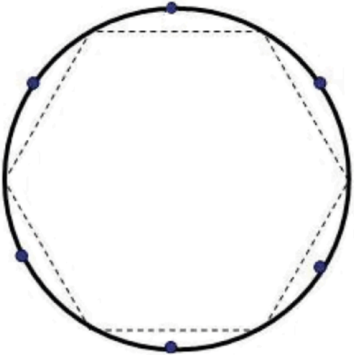
Examples of CVTs

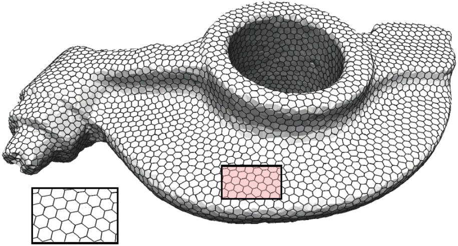


Constrained CVTs

For each Voronoi region V_i on surfaces S, the associated constrained mass centroid z_i^c is defined as the solution of the following problems:

$$\min_{z \in S} F_i(z), \quad \text{where} \quad F_i(z) = \int_{V_i} \rho(x)|x - z|^2 dx$$

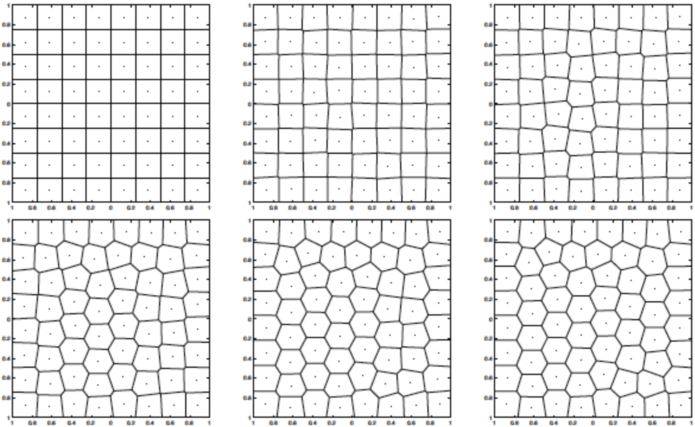




Courtesy of (Y. Liu, et al.)

Uniqueness of CVTs

- Starting from any CVT, you will end up with a CVT that is as hexagonal as possible



- Why hexagons? Because they possess a certain optimality property

Centroidal Voronoi Tessellations as minimizers

Given:

- $\Omega \subset \mathbb{R}^N$
- A positive integer k
- A density function $\rho(.)$ defined on $\bar{\Omega}$

Let

- $\{\mathbf{z}_i\}_{i=1}^k$ denote any set of k points belonging to $\bar{\Omega}$ and $\{V_i\}_{i=1}^k$ denote its corresponding Voronoi tessellation

Define the *energy functional*

$$\mathcal{G}\left(\{\mathbf{z}_i\}_{i=1}^k\right) = \sum_{i=1}^{k} \int_{V_i} \rho(\mathbf{y}) |\mathbf{y} - \mathbf{z}_i|^2 \, d\mathbf{y}.$$

The minimizer of $\mathcal{G}$ necessarily forms a CVT

Some Properties of CVTs

- If $\Omega \subset \mathbb{R}^N$ is bounded, then $\mathcal{G}$ has a global minimizer
- Assume that $\rho(.)$ is positive except on a set of measure zero in Ω
  - then $z_i \neq z_j$ for $i \neq j$
- For general metrics, existence is provided by the compactness of the Voronoi regions; uniqueness can also be attained under some assumptions, e.g., convexity, on the Voronoi regions and the metric

Gersho's conjecture

- For any density function, as the number of points increases, the distribution of CVT points becomes locally uniform
- In 2D, CVT Voronoi regions are always locally congruent regular hexagons
- In 3D, the basic cell of a CVT grid is truncated octahedron [Du/Wang, CAMWA, 2005]
- Gersho's conjecture is a key observation that helps explain the effectiveness of CVTs

Range of applications

- Location optimization:
  - optimal allocation of resources: mailboxes, bus stops, etc. in a city
  - distribution/manufacturing centers
- Grain/cell growth
- Crystal structure
- Territorial behavior of animals
- Numerical methods
  - finite volume methods for PDEs
  - Atmospheric and ocean modeling
- Data analysis:
  - image compression, computer graphics, sound denoting etc
  - clustering gene expression data, stock market data
- Engineering:
  - vector quantization etc
  - Statistics (k-means):
  - classification, minimum variance clustering
  - data mining

Optimal Distribution of Resources

What is the optimal placement of mailboxes in a given region?

- A user will use the mailbox nearest to their home
- The cost (to the user) of using a mailbox is proportional to the distance from the users home to the mailbox
- The total cost to users as a whole is measured by the distance to the nearest mailbox averaged over all users in the region
- The optimal placement of mailboxes is dened to be the one that minimizes the total cost to the users

Observation:

The optimal placement of the mail boxes is at the centroids of a centroidal Voronoi tessellation

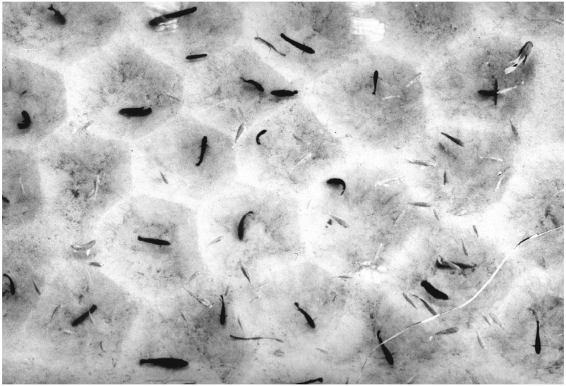
Cell Division

- There are many examples of cells that are polygonal often they can be identified with a Voronoi, indeed, a centroidal Voronoi tessellation.
  - this is especially evident in monolayered or columnar cells, e.g., as in the early development of a starsh (Asteria pectinifera)
- Cell Division
  - Start with a configuration of cells that, by observation, form a Voronoi tessellation (this is very commonly the case)
  - After the cells divide, what is the shape of the new cell arrangement?

Observation:

The new cell arrangement is closely approximated by a centroidal Voronoi tessellation

Territorial Behavior of Animals



A top view photograph, using a polarizing filter, of the territories of the male Tilapia mossambica

Photograph from: George Barlow; Hexagonal territories, Animal Behavior 22 1974, pp. 876878

Finite Volume Methods Having Optimal Truncation Errors

It has been proved that a finite volume scheme based on CVTs and its dual Delaunay grid is second-order accurate [Du/Ju, Siam J. Numer. Anal., 2005]

- this result holds for general, unstructured CVT grids
- this result also holds for finite volume schemes on the sphere

Lloyd's Method [Lloyd 1957]

1. Start with the initial set of points $\{z_i\}_{i=1}^K$
2. Construct the Voronoi tessellation $\{V_i\}_{i=1}^K$ of Ω associated with the points $\{z_i\}_{i=1}^K$
3. Construct the centers of mass of the Voronoi regions $\{V_i\}_{i=1}^K$ found in Step 2; take centroids as the new set of points $\{z_i\}_{i=1}^K$
4. Go back to Step 2. Repeat until some convergence criterion is satisfied

Note:Steps 2 and 3 can both be costly

Lloyd's method: analytical convergence results

Assumptions:
1) The domain $\Omega \subset \mathbb{R}^N$ is a convex and bounded set with the diameter

$$\mathrm{diam}(\Omega) := \sup_{\mathbf{z},\mathbf{y}\in\Omega} |\mathbf{z}-\mathbf{y}| = R_\Omega < +\infty.$$

2) The density function ρ belongs to $L^1(\Omega)$ and is positive almost everywhere. Consequently, we have that

$$0 < M(\Omega) = \|\rho\|_{L^1(\Omega)} = \int_\Omega \rho(\mathbf{y})d\mathbf{y} < +\infty.$$

- **Theorem 1.** The Lloyd map is continuous at any of the iterates.
- **Theorem 2.** Given $n \in \mathbb{N}$ and any initial point $\mathbf{Z}^0 \in \bar{\Omega}$. Let $\{\mathbf{Z}^i\}_{i=0}^{\infty}$ be the iterates of Lloyd algorithm starting with $\mathbf{Z}^0$. Then
  (1) $\{\mathbf{Z}^i\}_{i=0}^{\infty}$ is weakly convergent (i.e., $\lim_{i\to+\infty} \nabla\mathcal{G}(\mathbf{Z}^i) = 0$) and any limit point of $\{\mathbf{Z}^i\}_{i=0}^{\infty}$ is also a non-degenerate critical point of the quantization energy $\mathcal{G}$ (and thus a CVT).
  (2) Moreover, it also holds that $\lim_{i\to+\infty} \|\mathbf{Z}^{i+1} - \mathbf{Z}^i\| = 0$.

Du/E./Ju 2006, E./Ju/Rand 2008.

Accelerating convergence

Lloyd method (fixed-point iteration $z_{n+1} = Tz_n$) $\Rightarrow$ only linear convergence.

In 1D, for strongly log-concave densities the convergence rate of Lloyd's iteration was shown to satisfy

$$r \approx 1 - \frac{C}{k^2}$$

so the method significantly slows down for large values of k.

Empirical results show similar behavior for other densities.

Is speedup possible?

Newton-type and Multilevel method

Newton-type [Du/E., Num. Lin. Alg. 2006] :

$$\tilde{z} = z + \alpha(dT|_z - \mathbf{I})^{-1}(z - T(z))$$

This method was shown to converge quadratically for suitable initial guess.

Multilevel [Du/ E., SINUM 2006, 2008]:

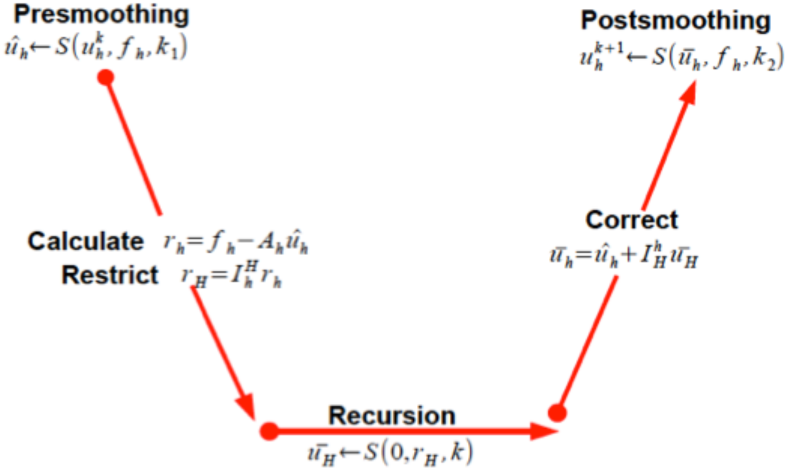
- Lloyd Method
- A spatial decomposition
- A multilevel successive subspace corrections

Illustration

(Loading CVT motion)

V-cycle Multigrid for $Au = f$

- Given:
  - an initial estimate u_h^0 of the solution u_h^* on the fine level
  - smoother $\bar{u} \leftarrow S(u_h, f_h, k)$



Other Extensions of Multigrid

- Full Approximation Scheme (FAS)
- Multigrid as a precondioner
- the multilevel adaptive technique (MLAT)
  - Exhibits adaptive mesh refinement
- Algebraic Multigrid (AMG)
  - Constructs the hierarchy level by only utilizing the information from the algebraic system to be solved
  - The purpose is to overcome the irregular underlying meshes

Multigrid in Optimization

- The necessary condition to solve $\min_z \frac{1}{2} z^T A z - f^T z$ is to solve $Az = f$
- The necessary condition to solve a more general optimization problem $\min_z f(z)$ is to solve a nonlinear system of equation $\nabla f(z) = 0$.
- One approach to solve systems of nonlinear equations is Full Approximation Scheme (FAS) which is a generalization of the traditional multigrid.
- In our approach, we choose multigrid-based optimization (MG/OPT) framework which relies explicitly on optimization models as subproblems on coarser grids.

Advantages of MG/OPT

- MG/OPT can deal with more general problems in an optimization perspective, in particular, it is able to handle inequality constraints in a natural way.
- MG/OPT has a better guarantees of convergence than using traditional multigrid for a system of equations.
- for a class of optimization problems governed by differential equations, multigrid will be better suited to the explicit optimization model rather than underlying differential equation when it is not elliptic.

Mutigrid Optimization framework (MG/OPT)

- Optimize a high-resolution model:

$$\begin{array}{ll} \textit{minimize} & f_h(z_h) \\ \textit{subject to} & a_h \leq 0 \end{array}$$

- An available easier-to-solve low-resolution model:

$$\begin{array}{ll} \textit{minimize} & f_H(z_H) \\ \textit{subject to} & \hat{a}_H \leq 0 \end{array}$$

MG/OPT components:

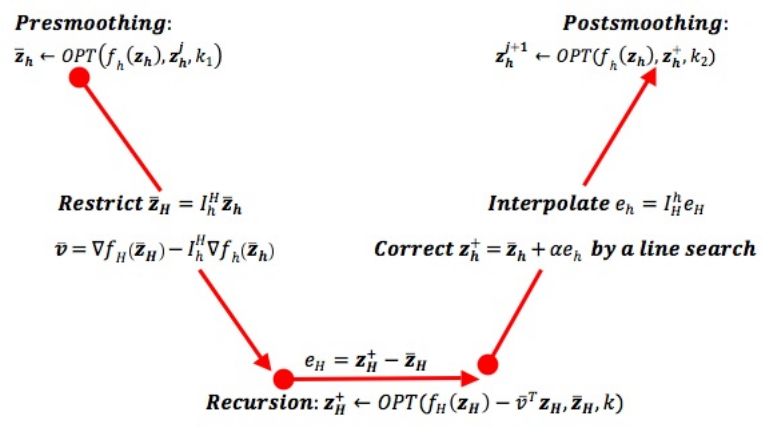
- Given $\min_z f(z)$ with initial guess $\bar{z}$, k -number of iterations, let OPT be a convergent optimization algorithm:

$$\mathbf{z}^+ \leftarrow \mathrm{OPT}(f(\mathbf{z}), \bar{\mathbf{z}}, k)$$

- h: fine grid; H: coarse grid
- a high-fidelity model f_h
- a low-fidelity model f_H
- a downdate operator I_h^H for the variables $\mathbf{z}_h$
- an update operator I_H^h

Multigrid Optimization Algorithm [S.G. Nash 2000] (MG/OPT: Unconstrained)

- Given:
  - an initial estimate $\mathbf{z}_h^0$ of the solution $\mathbf{z}_h^*$ on the fine level
  - Integers k_1 and k_2 satisfying $k_1 + k_2 > 0$



Multigrid Optimization Algorithm (MG/OPT: Constrained)

- On the fine level, solve:

$$\begin{array}{ll} \min_{z_h} & f_h(z_h) \\ \text{subject to} & a_h(z_h) \leq 0 \end{array}$$

- Set a downdate operator J_h^H for the Lagrangian Multipliers λ_h
- Construct the shifted model

$$\begin{aligned} \bar{\mathbf{z}}_H &= I_h^H \bar{\mathbf{z}}_h \\ \bar{\lambda}_H &= J_h^H \bar{\lambda}_h \\ \bar{\mathbf{v}} &= \nabla \mathcal{L}_H(\bar{\mathbf{z}}_H, \bar{\lambda}_H) - I_h^H \nabla \mathcal{L}_h(\bar{\mathbf{z}}_h, \bar{\lambda}_h) \\ f_s(\mathbf{z}_H) &= f_H(\mathbf{z}_H) - \bar{v}^T \mathbf{z}_H \\ \bar{s} &= a_H(\bar{\mathbf{z}}_H) - J_h^H a_h(\bar{\mathbf{z}}_h) \\ a_s(\mathbf{z}_H) &= a_H(\mathbf{z}_H) - \bar{s} \end{aligned}$$

- On the coarse level,

$$\begin{array}{ll} \min_{z_H} & f_s(z_H) \\ \text{subject to} & a_s(z_H) \leq 0 \end{array}$$

Convergence

- If
  - OPT is convergent
  - The objective function and constraints are continuously differentiable
- Then
  - MG/OPT converges in the same sense as OPT

Descent: Unconstrained version

The search direction e_h from the recursion step will be a descent direction for f_h at $z_{1,h}$ if

- $I_H^h = C(I_h^H)^T$
- $f_s(z_{2,H}) < f_s(z_{1,H})$
- $e_H^T \nabla^2 f_H(z_{1,H} + \alpha e_H) e_H > 0$, for $0 \leq \alpha \leq 1$

Descent: Equality Constrained version

The search direction e_h from the recursion step is guaranteed to be a descent direction for the merit function M_h at $z_{1,h}$ if

- $M_s(z_H^+) < M_s(\bar{z}_H)$
- $v^T \nabla^2 f_H(z_H) v$ for v in the null space of $\nabla a_H(x_H)$
- ρ is large enough so that $e_H^T \nabla^2 M_s(\bar{z}_H + \eta e_H) e_H > 0$ for $0 \leq \eta \leq 1$

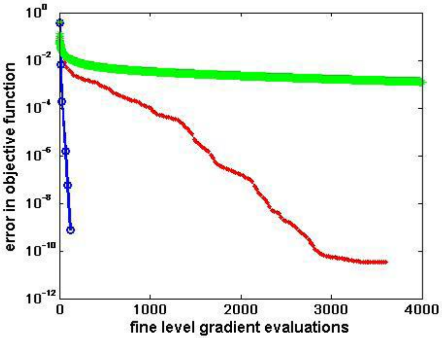
MG/OPT setup

- We have chosen OPT to be the truncated-Newton algorithm in our experiments
- 1-D case:
  - Downdate from finer to coarser grid:
    - Solution restriction (injection): $[I_h^H v_h]^i = v_h^{2i}, \quad i = 1, 2, \ldots, k/2$
    - Gradient restriction (scaled full weighting): $[\hat{I}_h^H v_h]^i = \frac{1}{2} v_h^{2i-1} + v_h^{2i} + \frac{1}{2} v_h^{2i+1}, \quad i = 1, 2, \ldots, k/2$
  - Update from coarser to finer grid: $I_H^h = (\hat{I}_h^H)^T$
- 2-D case:
  - Downdate from finer to coarser grid:
    - Solution restriction is performed by injection
    - Gradient restriction: $[\hat{I}_h^H v_h]^i = \sum_j \alpha_j^i v_h^j$, where $\alpha_i^i = 1$ and $\alpha_j^i = \frac{1}{2}$ for any j s.t. z_j is a fine node sharing an edge with z_i in the fine level triangulation
  - Update from coarser to finer grid: $I_H^h = 4(\hat{I}_h^H)^T$

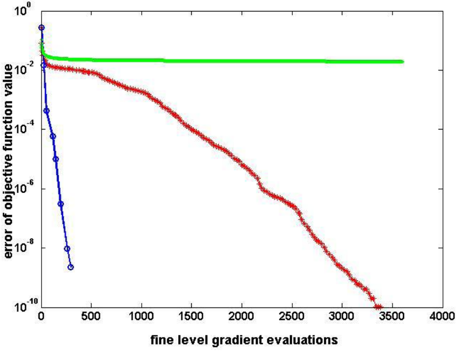
Convergence Result on 1-D CVT

Blue: MG/OPT; Red: OPT; Green: Lloyd

$\rho(y) = 1$

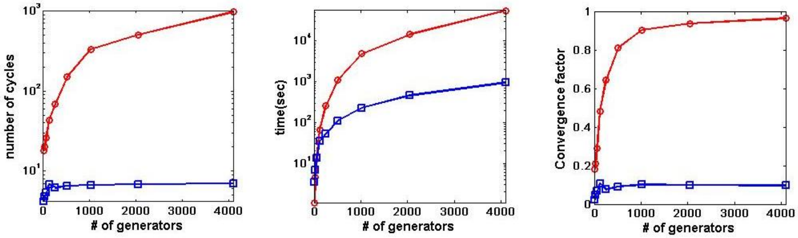


$\rho(y) = 6y^2 e^{-2y^3}$



Convergence Result on 1-D CVT (cont'd)

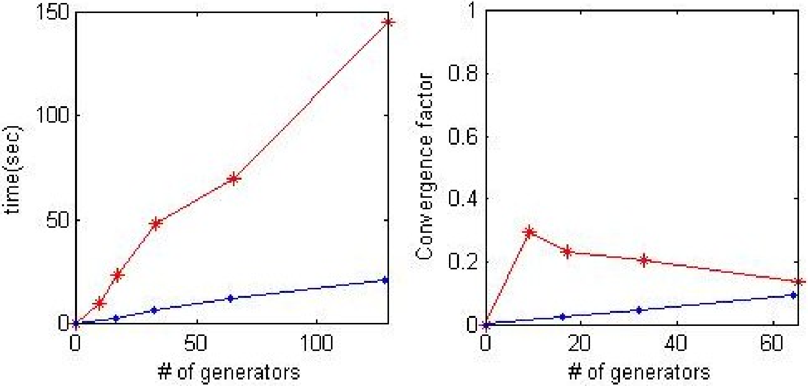
Solving problems of increasing size, MG/OPT versus OPT ($\rho(y) = 1$).
Blue: MG/Opt, Red: OPT;



Similar results for linear and nonlinear densities

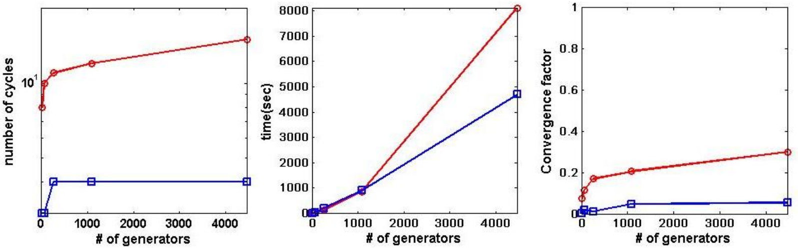
Comparison with Multilevel-Lloyd

Convergence factors for MG/OPT vs. Multilevel-Lloyd $\rho(y) = 1$; Blue: MG/OPT; Red: Multilevel-Lloyd



Convergence Result on 2-D CVT based on triangular domain

Red: Opt; Blue: MG/OPT; $\rho(x) = 1$



Di/Emelianenko/Nash, NMTMA, 2012

Results and challenges

- CVT is in the heart of many applications and the number is growing: computer science, physics, social sciences, biology, engineering ...
- Tremendous progress has been made in the last decade, but many questions remain unsolved, both theoretical and numerical.
- The main advantage of MG/OPT is its superior convergence speed, and the fact that it preserves low convergence factor regardless of the problem size.
- MG/OPT
- The simplicity of its design and the results of preliminary tests suggest that the method is generalizable to higher dimensions, which is the subject of further investigations

Future Work

- Develop MG/OPT for bound constrained setting with user requirements comparable to those for unconstrained setting.
- Implement MG/OPT for higher dimensional CVT problems with nontrivial densities and random initial configurations.
- Analyze the properties of the Hessian matrix for general CVT configurations.
- Future work also includes application of this technique to various scientific and engineering applications, including image analysis and grid generation.

Future Work

- Develop MG/OPT for bound constrained setting with user requirements comparable to those for unconstrained setting.
- Implement MG/OPT for higher dimensional CVT problems with nontrivial densities and random initial configurations.
- Analyze the properties of the Hessian matrix for general CVT configurations.
- Future work also includes application of this technique to various scientific and engineering applications, including image analysis and grid generation.

THANKS!