A SIMPLE GEOMETRIC METHOD FOR NAVIGATING THE ENERGY LANDSCAPE OF CENTROIDAL VORONOI TESSELLATIONS

IVAN GONZALEZ, RUSTUM CHOKSI, AND JEAN-CHRISTOPHE NAVE

Abstract. Finding optimal centroidal Voronoi tessellations (CVTs) of a 2D domain presents a paradigm for navigating an energy landscape whose desirable critical points have sufficiently small basins of attractions that they are inaccessible with Monte-Carlo initialized gradient descent methods. We present a simple deterministic method for efficiently navigating the energy landscape in order to access these low energy CVTs. The method has two parameters and is based upon each generator moving away from the closest neighbour by a certain distance. We give a statistical analysis of the performance of this hybrid method comparing with the results of a large number of runs for both Lloyd’s method and state of the art quasi-Newton methods.

Key words. Centroidal Voronoi tessellation, global optimization, energy ground state, optimal vector quantizer, regularity measures.

AMS subject classifications. 49Q10, 65D18, 68U05, 82B80

1. Introduction. A fundamental problem in information theory and discrete geometry is known, respectively, as optimal quantization and optimal centroidal Voronoi tessellations (CVTs), c.f [10, 11, 15]. For a bounded $\Omega \subseteq \mathbb{R}^d$, a collection of $N$ distinct points $\mathbf{X} := \{x_i\}_{i=1}^N \subset \Omega$ defines a tessellation $\mathcal{V}(\mathbf{X}) = \{V_i\}_i^N$ of $\Omega$ via the Voronoi cells

$$V_i(\mathbf{X}) = \{y \in \Omega \mid ||x_i - y|| < ||x_j - y||, \ \forall j \neq i\} \quad i = 1, ..., N$$

For any $\mathbf{X}$, we define the following non-local energy:

$$F(\mathbf{X}) = \int_\Omega \text{dist}^2(y, \mathbf{X}) \, dy = \sum_{i=1}^N \int_{V_i} ||y - x_i||^2 \, dy =: \sum_{i=1}^N F_i$$

Criticality of this energy gives rise to a CVT; that is, finding a placement of the generators $x_i$ such that they are exactly the centroids of their associated Voronoi cell $V_i$, i.e.

$$x_i = c_i := \frac{1}{|V_i|} \int_{V_i} y \, dy \quad i = 1, ..., N.$$ 

In the context of information theory, the set $\mathbf{X}$ is viewed as a quantizer to quantize data; here, uniformly distributed over $\Omega$. The quantization error is given by $F(\mathbf{X})$ with the optimal quantizer the one with least error (alternatively the CVT with lowest energy). The energy (1.2) has a wealth of critical points (CVTs) [27, 40] and low energy CVTs have tiny basins of attractions making them difficult or impossible to find via gradient based descent with random initializations. On the other hand, one does have a benchmark for the optimal geometry in the limit $N \to \infty$ wherein we dispense with shape and boundary effects. Indeed, Gersho’s conjecture [14] addresses the periodic structure of the optimal quantizer as $N \to \infty$. The conjecture is completely
solved in 2D wherein the optimal Voronoi cell is the regular hexagon, corresponding to generators on a triangular lattice. However, to date it remains open in 3D wherein the belief is that the optimal Voronoi period-cell is the truncated octahedron, corresponding to generators on a body centred cubic (BCC) lattice; see [1, 5, 12].

The purpose of this article is to present and assess in 2D a simple deterministic method for efficiently navigating the energy landscape in order to access low energy CVTs which are otherwise inaccessible with Monte Carlo initializations coupled to gradient based descent methods (i.e. taking optimal results amongst hundreds of thousands to millions of gradient based descents on randomly sampled initial configurations). The proposed method has two parameters and is based upon each generator moving away from the closest neighbour (MACN) by a certain distance. There are two choices of distance, each with a different role:

- **(An iterated preconditioning step)** The distance to the centroid of the generator’s Voronoi cell. We refer to this step as $MACN-c$.

- **(A single geometric-based "annealing" step)** The fixed distance

  \[
  \delta := \frac{1}{4} \sqrt{\frac{|\Omega|}{N}},
  \]

  a parameter set in terms of the average distance between generators in the approximate regular hexagonal lattice. We refer to this step as $MACN-\delta$.

Our method is then an iterative procedure starting with a random initialization (placement of generators):

Step 1 Iterating the $MACN-c$ preconditioning step $K$ times;
Step 2 Lloyd’s or any other deterministic method surveyed in §3 that minimizes (1.2) to a CVT;
Step 3 "Annealing" once with $MACN-\delta$;
Step 4 Repetition of Steps 1-3 a number of $Q$ times, ending at Step 2.
Step 1 does result in a configuration which is often, yet not necessarily, close to a CVT. However, this is not the point: It results in a configuration which lies in the basin of an energetically desirable CVT. By energetically desirable we mean two things: (i) it has low energy in the sense that its energy is comparable with the optimal result of any standard gradient-based descent algorithm assessed over a "large" number of runs (for us with $N \sim 1000$, a large number of runs is on the order of $100,000$); (ii) the same holds true for the measures of regularity described in §2. Step 2 achieves this energetically desirable CVT. Step 3 breaks away from this basin to another basin which can contain a more optimal CVT. The choice of $\delta$ in (1.3) is subtle: it is large enough in order to change basins but sufficiently small in order to not loose the desired regularity achieved thus far. Note that in 2D the optimal configuration likes to settle on $N$ regular hexagon Voronoi cells. In this case, $\delta$ can be thought of as half the distance from the generator to its Voronoi cell boundary plane. Indeed, in the case of regular hexagonal Voronoi cells, the 6 neighbouring generators are equally distant. In this scenario, Step 3 chooses one of the 6 closest generators according to some pre-established rule, possibly a random selection linking with a more familiar simulated annealing mechanism.

We implement and assess our method on two choices of 2D domains, both having periodic boundary conditions in order to dispense with boundary effects. To start we work on a primary domain for which the ground state is known: the periodic regular hexagon which can be tessellated into $N$ regular hexagons provided $N$ is suitably chosen [6]. We then work on a primary domain that does not permit a perfect regular tilling: the flat square torus. Here there is always frustration due to the size effects, and it is surprisingly unclear as to the true nature of the lowest energy state. With the number of generators $N$ taken from the range 1000 to 4000, we show that our hybrid method with $Q \leq 10$, implemented with less than two dozens initializations, readily finds states with far lower energy (and other metrics of optimal regularity) than the ones accessible with Lloyds' method, or any state of the art gradient based descent method, assessed over hundreds of thousands random initializations, see Figure 1.2. In §5 we present the full details of this comparison, emphasizing the role of statistics for assessing ours and other numerical methods.
The motivation and scope of our hybrid method is twofold. First, there is the direct application to generating low energy and regular CVTs. In contrast to methods which are based upon initial sampling or building on regular hexagons, our method is based only on basic structures of the energy –centroids and distance functions–. Moreover, on domains like the square torus, optimality is subtle with regard to the presence of non-regular hexagons and defects (non-hexagonal cells): indeed, there is probably not even one perfectly regular hexagon in the Voronoi diagram of the optimal energy configuration.

There is a second, more general, scope to our work. Probing non-convex and non-local energy landscapes is a fundamental problem in physics and applied mathematics. The CVT energy (1.2) is perhaps the simplest non-trivial example; while it is finite dimensional with a simple geometric characterization of criticality (namely a CVT), it is challenging (even in 2D) to navigate its landscape (the landscape of CVTs). Our results demonstrate that while the CVT energy landscape on the square torus with \( N \sim 1000 \) is indeed complex, our hybrid algorithm is able to efficiently navigate it with only 9-10 deterministic "annealing" steps.

Finally let us remark on the empirical nature of our work. While we give a heuristic rationale for our MACN steps, the precise nature of the distance chosen is based in part on empirical tries. Indeed, for the preconditioning step, we tried a variety of other movements; for example, movement towards the furthest neighbour, MACN individually tuned to the distance to the closest generator. None of these performed as well as the centroidal distance. For the MACN-\( \delta \) step, we experimented with other choices of \( \delta \), for example the intrinsic length-scale of the Voronoi cell (cf. the Appendix A). Our choice is one with sensibly the most effective performance. Overall, while we only have heuristics to explain certain aspects of our method, we feel the remarkable results justify its presentation and discussion here.

The paper is organized as follows: in §2 we give a brief description of periodic centroidal Voronoi tessellations (PCVTs). We also discuss, besides the normalized energy, certain natural local measures of regularity, one of which is novel. Then in §3,
we survey methods to generate and improve CVTs. We present our hybrid algorithm in §4 and then an analysis for its predictions in §5. We finish with closing remarks and future directions in §6.

2. Periodic CVTs and Regularity Measures. For the remaining of the paper we will consider 2D torii spaces; namely polygons Ω, with opposite and pairwise identifiable sides, that can be periodically extended to the plane. With the enforcement of these boundary conditions follow the definitions of a periodic-VT (PVT) and periodic-CVT (PCVT) by using the metric $||.||_T$ that is inherited from the Euclidian metric in the above definitions (1.1) and (1.2) respectively. See Figure 1.1 (c) and (d) for a concrete example.

In addition, the dual graph of the set $\cup_{i\leq N} \partial V_i$ known as the periodic-Delaunay triangulation (PDT) will be of importance. The PDT has an edge $e_{ij}$ linking $x_i$ and $x_j$ iff they are neighbours in the PVD, in other words the index sets $\mathcal{N}_i := \{ j \neq i \} \cap \partial V_j \cap \partial V_i \neq \emptyset$; $i = 1,...,N$ are intrinsic to the triangulation.

We refer the reader to [34] for an introductory treatment on Delaunay Triangulations and other variants of the Voronoi Tessellation and to [40, 39, 4] for a detailed definition of the periodic framework.

Centroidal characterizations. In addition to the geometrical property $x_i = c_i \ \forall i$, a PCVT admits a variational characterization via $F : \mathcal{D} \to \mathbb{R}_+$ given by (1.2), here $\mathcal{D} := \{ X \in \mathbb{R}^{2N} | x_i \neq x_j \ \forall j \neq i ; \ x_i \in \Omega \ \forall i \}$ represents the set of non-degenerate configurations of generators in our periodic space. Depending on the application the functional $F = \sum_i^N f_i$ is often referred to as the distortion, cost function or potential. Alternatively, one can interpret each $f_i$ from a physical point of view as the trace of the $2 \times 2$ inertia tensor of $V_i$ with respect to $x_i$, i.e the resistance of $V_i$ to its rotation around an axis passing through $x_i$ that is orthogonal to $\Omega$.

It is a well established fact in rigid mechanics (parallel axis theorem) that the trace of this tensor will be minimized whenever the orthogonal axis of rotation passes through the centroid. This locally translates to a Voronoi region being generated by $x_i$ through the centroid. This locally translates to a Voronoi region being generated by $x_i$.

Furthermore $F$ has been shown to be $C^2(D)$ [25, 40] and although the Hessian is
sparse it is in general not definite, as noted in [27] this is a consequence of the energy functional being highly non-convex with the presence of saddle points, see also [40].

**Optimal PCVT, ground state energy and regularity measures.** As for the majority of non-convex problems, finding the global minimum of $F(X)$ and a corresponding global minimizer (ground state) remains an open question. Yet, to better assess our performance in the matter of reaching minimizers close to the ground state we introduce, using the established hexagon theorem [16, 31, 37] and in a similar spirit to [7, 38], an energy functional scaled with respect to $N F_{hex}$.

Here $F_{hex}$ denotes the second moment of a regular hexagon with area $|\Omega|/N$ and a simple calculation shows that

$$F_{hex} = \frac{5}{18\sqrt{3}} \frac{|\Omega|^2}{N^2}.$$  

The scaled energy that is thus independent of the size of the problem is

$$E(X) := \frac{F(X)}{NF_{hex}} = \frac{1}{N} \sum_{i=1}^{N} \frac{F_i}{F_{hex}} =: \frac{1}{N} \sum_{i=1}^{N} E_i$$

and with the same scaling carrying out trivially to $DE(X)$ and $D^2E(X)$.

As an additional way to help us quantify the quality of a PCVT, and as it has been customary done in the literature, we use the fraction of hexagonal cells:

$$H(X) := \frac{\{V_i, i = 1, \ldots, N | \#V_i = 6\}}{N}$$

However, regularity measures such as $H$ or the Voronoi Entropy used in [3] (and rediscussed in [2]) only give us information about the connectivity of the DT and make no direct connection with the hexagon theorem requiring hexagons to be regular to achieve the energy $E = 1$. For this reason we create a refinement on $H$ that takes into account the regularity of the hexagonal cells through their isoperimetric ratio. We recall that the isoperimetric ratio $r$ of a polygon is the dimensionless ratio of its perimeter squared and area. Using $r(V_i)$ of each Voronoi cell together with the isoperimetric ratio of the regular hexagon $r_{hex} := 8\sqrt{3}$, and for given $\epsilon > 0$ small we introduce the following regularity measure of a tessellation $\mathcal{V}(X)$:

$$R_\epsilon(X) := \frac{\{V_i, i = 1, \ldots, N | \#V_i = 6 ; \frac{|V_i|}{r_{hex}} \leq \epsilon\}}{N}.$$

That is $R_\epsilon$ is the fraction of cells that are regular hexagons within an isoperimetric tolerance of $\epsilon$. Generators whose cell is taken into account by $R_\epsilon$ will be depicted in red in the PVDs of Figure 4.1 and onwards.

Later in §4 and §5 we will rely on $E - 1$, $H$ and $R_\epsilon$ to measure the performance of our method and others in reaching high quality PCVTs. The data will also make the compelling case that $R_\epsilon$ is a more sensitive measure of low values of $E$ than $H$ (for appropriate $\epsilon$). Finally the statistics on these three quantities will be at the core of the conjecture that our method acts at a semi-global scale.
3. Generating and improving CVTs. We divide this brief literature survey into methods available to compute CVTs and ways to enhance their quality.

Computing CVTs. These methods rely either on the characterisation $x_i = c_i \forall i$ and seek to solve these nonlinear equations or are of variational character on $E(X)$:

1. Lloyd’s method, introduced in the seminal work [26], is unquestionably the most widespread method due to its simplicity. It iteratively applies the map $T : D \rightarrow D$ defined by $T(X) = C$ where $C := \{c_i\}_{i=1}^N$. We refer the reader to [10, 9, 13] for properties and analysis of the map.
   Due to the importance of the method in the remaining of this paper we provide explicit pseudo-code as a sub-routine in Appendix B.

2. McQueen’s and probabilistic Lloyd’s [23], generate small sets of random sampling points in $\Omega$ and use these to approximate the centroid of Voronoï regions.

3. Lloyd-Newton [8, 25] uses Newton’s root finding method on $S(X) := X - T(X)$ after some Lloyd steps to reach the vicinity of a root in $D$. However the method may produce unstable CVTs (saddle points of $E$) if the number of Lloyd steps is not big enough.

4. Newton’s classical technique that minimizes Hessian-based quadratic models of $E \in C^2(D)$. With (2.1) available, the method converges at least quadratically when coupled with a line-search ensuring the strong Wolfe conditions [33]. Nonetheless this method suffers from two downsides: i) $D^2E$ is often indefinite and needs to be altered, for example by adding a “small” matrix to render it SPD prior to executing an incomplete Cholesky factorization, see [28, 24] for theory and low memory algorithms. ii) the Hessian is expensive to populate due to the boundary integrals.

5. Quasi-Newton BFGS collection. These methods only use $E$ and $DE$ to give an iterative approximation of the inverse Hessian. They remain to this day the favoured methods in the literature for fast CVT/PCVT computation due to their expected super-linear convergence whenever they are coupled with a line-search method that ensures the strong Wolfe conditions. The two families suited for medium/large scale problems that will be used in this paper are:
   
   (a) the low memory L-BFGS($M$) in which the inverse Hessian approximation remains sparse and is computed recursively from the $M$ previous approximations.
   
   (b) the Preconditioned-L-BFGS($M,T$) uses, every modulo $T$ iterations, a SPD preconditioner matrix $\tilde{A}$ (that does not necessarily need to approximate the Hessian) with the goal of redirecting the algorithm to a more suitable descent direction.

See [33] for a thorough description and analysis of the classical BFGS and L-BFGS($M$) and [22, 25] for the use P-L-BFGS($M,T$). Finally, an explicit routine of the preconditioned algorithm is provided in Appendix C.
6. Non-Linear Conjugate Gradient (NLCG) methods generalize the classical CG used in quadratic programming. Several updates for the conjugate directions are available [17, 33]. One can also use a relevant preconditioner SPD matrix \( \tilde{A} \) to improve the descent-conjugate directions.

There are several preconditioner matrices \( \tilde{A} \), for example: the Hessian \( D^2E \) itself (along with the often necessary modification to make it SPD) or a Graph Laplacian \( G \) introduced in [20] whose purpose is to approximate \( DE(X) = 0 \) to first order by the matrix equation \( GX = 0 \).

Originally presented for the \( \Omega \)-bounded case, below is our adaptation of \( G \) for the periodic case; denoting by \( \tau_{i,j} \) the pyramid with base \( \partial V_j \cap \partial V_i \) and apex \( x_i \) then the \( N \times N \) matrix is given by

\[
G := \begin{cases} 
  g_{ij} = -\int_{\tau_{i,j} \cup \tau_{j,i}} \rho(y) dy & \text{if } j \in N_i \\
  g_{ii} = \sum_{j \in N_i} |g_{ij}| & \text{otherwise} \\
  0 & \text{otherwise}
\end{cases}
\]

However, contrary to the original construction for \( \Omega \)-bounded, our adaptation is symmetric and positive semi-definite (\( G \) given by (3.1) is not strictly diagonal dominant and can thus be singular). As a consequence we will need to use a modified Cholesky factorization of \( G \).

**Energetic improvements.** Next we survey initialisations and other methods used jointly with the above algorithms in order to improve the quality of CVTs.

**Initialisations.**

(a) A Greedy Edge-Collapsing initialization [30] that meshes \( \Omega \) using more than \( N \) vertices, then it repeatedly uses an edge-collapsing scheme and finally the decimated vertices are employed as the initial generators.

(b) Quasi Random samplings of \( \Omega \) reduce the discrepancy of the initial cloud of generators, i.e the sampling of each site depends on the position of the others. These QR samplings use low discrepancy sequences such as Halton’s, Hammersley’s, Niederreiter’s and Sobol’s [32, 19, 18], see also [35] for CVT results.

**Couplings.**

(a) A Hierarchical method [38] that refines a CVT by cleverly inserting new generators over the DT. In this way the "regular" portions of the CVT "grow" when alternating with an energy descent method.

(b) An Atomic Operation method [29], here the authors establish three operations on “defects” that merge or split non-hexagonal cells prior to minimizing the energy; the process is then repeated.

(c) A Global Monte Carlo method [27] which applies ideas from Simulated Annealing. This method starts at a CVT and after a specific random perturbation of generators, that is dependent on the size of each \( V_i \), a new CVT is obtained by a non-linear minimization method. If the new energy is lower than the previous one then the algorithm automatically accepts that new
configuration, otherwise it accepts it only according to a transitional probability that is dependent both on a cooling temperature and on the energy gap between the two CVTs. A crucial disadvantage is the number of parameters that need to be adjusted to obtain good performances, namely: the initial temperature, the temperature decay to zero, the perturbation amplitude and the number of iterations repeating the procedure.

The above-mentioned literature shows that these initialisations and couplings yield significant energetic improvements, even more so when combined; for this reason they find use in many applications such as optimal meshing and computer graphics. Nonetheless, the initialisation procedures lacking dynamics and the couplings (a) and (b) above that yield a number of generators other than $N$ are not suitable to tackle another variety of problems such as optimal quantization and geographical optimization.

4. Our hybrid algorithm. We divide this section in two parts: first we present our algorithm’s first stage ($MACN-c$) which can be regarded as a dynamical initialization, similar in outcome to those presented in §3, that we later couple with Lloyd’s method to reach a PCVT. Second, we construct subsequent stages from criticality in order to obtain a sequence of potentially lower energy states.

$MACN-c$. We start with a collection $X(0)$ of initial generators. Then for a given input parameter $K_0$ representing the number of iterations used to get the initialization $X(K_0)$, our update scheme is

$$x_i \leftarrow x_i + \|x_i - c_i\| \frac{x_i - x_{j_i^*}}{\|x_i - x_{j_i^*}\|}, \quad i = 1, \ldots, N$$

where $x_{j_i^*}$ is a closest Voronoi neighbor to $x_i$ at the current iterate, i.e $x_{j_i^*}$ solves $\min_{j \neq i} \|x_i - x_j\| = \min_{j \in \mathcal{N}_i} \|x_i - x_j\|$ where $\mathcal{N}_i$ was defined earlier to be the index set of Delaunay edge-connected neighbors to the site $x_i$. Notice that $x_{j_i^*}$ may not be unique however it suffices for our implementation to simply pick one solution via a pre-established tie breaking rule (e.g. selection of the closest candidate up to machine precision).

The immediate advantage of having the $MACN-c$ dynamics is that it allows us to obtain a more evenly distributed set of generators, thus having a similar purpose to other initialization methods. Indeed when applying $MACN-c$ we obtain a significant decrease in energy, often close to two orders of magnitude, when compared to a general random sampling, see Figure 4.1. We remark that, excluding the degenerate case where generators coincide, the system governed by (4.1) presents a fixed point iff it reaches a PCVT. However the scheme is in general not contractive due to the abrupt topological changes in the PDT from one iteration to another. Thus these iterates will almost always fail to yield a PCVT, even when $K_0 \to \infty$. Hence, we implement Lloyd’s algorithm on $X(K_0)$ to recover $X_0^*$, a stable local minimizer of $E$, along with its corresponding PDT and PCVT.

Concerning the complexity and time efficiency of our initialization: most of the
Figure 4.1. Example of MACN-c dynamics with $N = 1500$ in a square $\Omega$. The $k^{th}$ iterations displayed start from the uncorrelated uniform random sampling of the domain shown in (a), the table contains the measures of the respective tessellations. The colouring represents $E_i$ and generators in red are those taken into account by $\mathcal{R}_\epsilon$ ($\epsilon = 0.5\%$). Finally we plot the three regularity measures against the iterate number up to $k = 7 \times 10^4$ to show the different regimes; the system’s meso-stability after $k \approx 10^4$ is apparent.

Software used to construct VT/PVT with Euclidean distance (e.g., the CGAL Computational Geometry Algorithms Library –https://www.cgal.org– for C++ or the built-in MATLAB function `voronoin`) rely on an early construction of the Delaunay Triangulation. Thus one can extract $\mathcal{M}_i \forall i$ before the computation of the tessellation
while keeping the same complexity. Moreover, the remaining difference between one step of \( MACN-c \) and one Lloyd step is the computation of the indices \( j^*_i \forall i \) which simply adds a lower order term \( O(N) \) to the optimal overall complexity \( O(N \ln(N)) \) in 2D \([34]\). Hence, our initialization scheme (4.1) benefits from the same low complexity of Lloyd’s algorithm as well as a comparable simple implementation. This is the reason we chose Lloyd’s for our coupling rather than any other gradient based method.

Finally we discuss the choice for the parameter \( K_0 \). We emphasize that to some extent it is in the value of \( K_0 \) that the trade off between time and energy performance is made; in general the \( MACN-c \) dynamics reach a "low" energy plateau and oscillate around it, c.f Figure 4.1. Thus for optimal results \( X(K_0) \) should attain that regime. However it is difficult to estimate a priori when the system will reach such a mesostability and even more so the required \( K_0 \) might be too large. Thus in practice \( K_0 \) should be chosen by the user to retain tractability.

**\( MACN-\delta \) and further stages.** We now expand the method with the goal of probing local minimizers of \( E \) that are in the vicinity of \( X_0^* \) that was found by combining \( MACN-c \) with Lloyd. The construction is as follow:

First we perturb \( X_0^* \) with a \( MACN-\delta \) step using \( \delta = (1/4) \sqrt{|\Omega|}/N \) as in (1.3), i.e we perform the update

\[
x_i \leftarrow x_i + \delta \frac{x_i - x_{j_i^*}}{||x_i - x_{j_i^*}||} \quad i = 1, ..., N
\]

See Figure 4.2 for a concrete example. After this annealing step we implement \( K_1 \) iterations of \( MACN-c \) followed by Lloyd’s again to obtain a new minimizer \( X_1^* \) and its corresponding PDT/PCVT.

In the hopes of trying to elucidate this a priori mysterious factor \( 1/4 \) perhaps the first ascertainment is that \( \delta \) equals the intrinsic length-scale of the regular hexagonal tessellation, namely \(|V|/|\partial V|\) where \( V \) represents here the regular hexagon with inscribed circle of diameter \( \sqrt{|\Omega|}/N \). More details on the global behaviour of \( MACN-\delta \) type perturbations are provided in Appendix A.

![Figure 4.2](image-url). Example of a MACN-\( \delta \) perturbation with \( N = 1000 \) in a square \( \Omega \). On the left is a generic PCVT and on the right is its resulting perturbation with \( \delta = \frac{1}{4} \sqrt{|\Omega|}/N \). The colormap represents energy and hexagonal cells with red generators are regular within the isoperimetric tolerance \( \epsilon = 0.5\% \).
We may further expand the method by repeating this process a total of \( Q \) times. For generality, one can define an integer sequence \( \{ K_q \}_{q=0}^{Q-1} \) for the number of MACN-c iterations performed during each of the \( Q \) blocks. Similarly to our previous discussion on the value of \( K_0 \), this sequence should be chosen such that \( \sum_q K_q \) maintains tractability. The resulting extended algorithm with \( Q \) stages is explicitly described in Appendix B for completeness.

We argued earlier that the reduction of discrepancy of our initialisation is common in outcome to other initialisations recalled in §3 and thus the success of our first stage MACN-c+Lloyd in reaching "low" energy PCVTs is not surprising. Then by introducing the \( Q-1 \) supplementary stages we created a coupling technique, i.e our alternating creates a symbiotic system of relaxation and contraction in which our MACN blocks help Lloyd’s algorithm in reaching lower energies states at the same time that Lloyd’s method helps us finding basins of attraction around which the combination of our perturbation and "initialization" explore the landscape. This remains in essence very similar to what is accomplish by other coupling methods such as the Global Monte Carlo approach.

However, the remarkable novelty and change of paradigm here, is that we probe the energy in a dynamical and completely deterministic way while systematically preserving the number \( N \) and with a method that is simply motivated by the geometry of the problem.

Finally, we emphasize that the sequence \( \{ E(X_q^*) \}_{q=0}^{Q-1} \) obtained with our method may not be strictly decreasing. It is in fact quite likely that our algorithm moves to a higher energy basin of attractions from one stage to another, thus resembling Simulated Annealing in that sense (whilst remaining completely deterministic). The advantage however is that with a reasonably low probing number \( Q \leq 10 \), we are able to sample low energy states that are impossible or scarcely achievable by other deterministic methods (even when using a stochastic sampling of the landscape done with a great number of initial configurations).

5. Numerical results. A total of six Examples with different \( N \)’s are presented below: the first two are on the regular hexagonal torus while the remaining four are on the square torus. In the former the perfect honey comb lattice is attainable iff \( N = a^2 + ab + b^2 \quad \forall a, b \in \mathbb{N} \) [6] while in the latter the ground state is unknown due to size frustration.

We reduce our method to having one parameter by fixing the probing number \( Q = 10 \) and imposing \( K_q \equiv K \quad \forall q = 0, ..., Q-1 \). Our study of the six cases thus begins by finding a constant \( K \) offering a suitable trade-off between time and energy. Once \( K \) is chosen we deepen the analysis of the performance of our method and then finish the section with a comparison of \( R_\epsilon \) and \( H \) as measures faithful to \( E \).

Throughout the section we use uniformly uncorrelated initial configurations over \( \Omega \) and our hybrid method is compared with L-BFGS(7) and P-L-BFGS(20,20), for the latter our implementation of the pseudo code of Appendix C uses the periodic adaptation of the preconditioner matrix \( \tilde{A} = G \) as described in (3.1). Additionally, we include results obtained with Lloyd’s in our periodic set up to have a solid point of reference for past and future work since this is the only algorithm with a complete lack of tuning parameters.
Following collectively the results of [25, 38, 20] as well as our own implementation
on some of the deterministic methods recalled in §3, we believe that the two Quasi-
Newton choices of comparison paint a good overview of the current deterministic state
of the art methods: in particular, \( \tilde{A} = D^2E \) as well as other \( (M,T) \) values were tested
for the torus but did not achieve noticeable systematic improvements in the regularity
measures.
We further emphasize that, while the objective and main contribution of this paper is
to establish a dynamical and fully deterministic way of sampling energy basins with
as fewer parameters as possible, we tested the Global MCM method on the torus with
parameter values provided in [27]; we report that the energy results are comparable
to ours, however the probing number for MCM to reach the lowest energies statisti-
cally sampled with Quasi-Newton methods is in the hundreds whilst ours in the few
dozens or less. Furthermore, although not adapted to the periodic boundaries, quasi
random samplings of \( \Omega \) were also tested prior to Quasi-Newton minimization; the
energies achieved out of 1,000 runs were comparable with those of the basins sampled
by Lloyd’s out the 10,000 runs initialized with uncorrelated distributions that are
discussed below.

Let us next introduce some notation; with \( K \) fixed our hybrid method and its
lowest sampled energy \( E_{\text{min}} \) will primarily be compared with respect to the reference
energy \( E_{\text{ref}} \) given by the minimal energy PCVT obtained from a large batch of initial
configurations using the three gradient based methods. More precisely, \( E_{\text{min}} \) will be
recorded with our method over 100 or 1,000 initial configurations (depending on the
Example) while \( E_{\text{ref}} \) will be the lowest energy amongst 100,000 runs for each of L-
BFGS(7) and P-L-BFGS(20,20) as well as 10,000 runs for Lloyd’s.
We define then the following performance ratio for the sampled minimal energies.

\[
\tau := \frac{E_{\text{min}} - 1}{E_{\text{ref}} - 1}
\]

We will also employ the empirical cumulative distribution functions (ECDFs)
\( f_{E-1}, f_{R}, f_{H} \) of our respective regularity measures as well as the values \( \tilde{f}_{E-1}, \tilde{f}_{R}, \tilde{f}_{H} \)
obtained when evaluating the ECDFs of our hybrid method at \( E_{\text{ref}}, R_{\text{ref}}, H_{\text{ref}} \)
respectively; these quantities will establish the frequency of PCVTs for which our hybrid
algorithm outperforms the best comparative method. Other basic statistics provided
on regularity measures include averages and standard deviations taken over the desig-
nated number of runs, we denote them by \( \langle \cdot \rangle \) and \( \sigma(\cdot) \) respectively. Finally we fix the
isoperimetric tolerance to be \( \epsilon \equiv 0.5\% \), further detail on this choice will be discussed
at the end of the section.

A final note on our energy measurements; no quadrature was involved (i.e exact
calculations were performed) and a tolerance was used on \( ||DE||/N \) guaranteeing that
the values of energy listed in all the tables and figures are accurate at least up to the
significant digits provided.

**Choosing \( K \).** The two scenarios on the hexagon \( \Omega \) (allowing the honey comb
tiling) are with \( N = 973 = 17^2 + 17 \times 19 + 19^2 \) and \( N = 2029 = 25^2 + 25 \times 27 + 27^2 \).
On the square we’ll work with the values \( N = \{n \times 1000\}_{n=1}^4 \). For these set ups we
run our hybrid method on a reduced set of 15 initial configurations using a selected
list of \( K \)’s, the energy results shown in Figure 5.1 will allow us to choose trade-off
values between time and energy. We remark on the general decrease tendency over the \( Q \) hybrid stages but that the decrease is not monotone in \( K \).

Note as well that because the first two values of \( N \) on the square are close to those on the hexagon there is no need to run sweeps for \( N = 1000 \) and 2000, we’ll just retain the same parameter values.

The graphs of \( \langle E - 1 \rangle \) in Figure 5.1 (a) and (b) suggest that we pick \( K = 6000 \) for Examples 1 & 3 and \( K = 8000 \) for Examples 2 & 4. For Examples 5 & 6 however, energy averages do not provide clear insight, we turn then to minimums from which Figure 5.1 (c) and (d) suggest we pick \( K = 8000 \) and \( K = 12000 \) respectively.

![Graphs](image)

**Figure 5.1.** \( K \)-sweeps: the joint markers represent \( \langle E - 1 \rangle \) while the isolated ones represent minimums over 15 initial configurations. In black dotted lines are the values \( E_{\text{ref}} - 1 \) appearing in Tables 1, 2, 5 and 6 for a first comparison.

**Hexagonal torus \( \Omega \).**

**Example 1.** With \( N = 973 \) and \( K = 6000 \) chosen, we run our method on a larger batch of 100 initial configurations; it reached the ground state with \( E_{\text{min}} - 1 \approx 1e^{-14} \) (maximal precision allowed by our implementation) while the optimal PCVT from the comparative methods is \( E_{\text{ref}} - 1 = 0.00287 \) (achieved by L-BFGS(7)), see Figures 5.2 & 5.3 as well as the statistics summary of Table 1. In particular, Figure 5.2 (c) shows how the symbiotic blocks of MACN-\( \delta \) + MACN-\( c \) act on probing non-PCVT configurations with energy close to (if not below) \( E_{\text{ref}} \).

Notice at last how the values \( f_{E-1}^{*}, f_{R}^{*}, f_{H}^{*} \) indicate that our method outperforms the others for a significant fraction of the runs. In particular, not only do we get the ratio \( \tau \) (5.1) to be sensibly zero but; on average one needs to run our hybrid algorithm on approximatively three uniformly sampled initial configurations to obtain an energy lower that \( E_{\text{ref}} \). In other words, we only need \( \approx 30 \) PCVTs so that our way of probing the energy landscape achieves comparable results to the sampling of the basins done by L-BFGS(7) out of 100,000 runs.
Example 2. For $N = 2029$ we first remark from the reduced sets of runs from Figure 5.1 (b) that energy averages and their deviation from $E^{ref}$ immediately compare favourably to those in Example 1, this is the first indicator that our hybrid is less affected than the comparative methods by the increasing non-convexity of the energy with $N$. The same conclusion can be drawn from the data from a hundred runs in Tables 1 and 2; more precisely we have that although our hybrid did not achieve the honey comb structure, our optimal PCVT with $E^{min} - 1 = 0.00150$ is still far more regular and gets $\tau^{-1} \approx 3.2$ times closer to the ground state than the compared methods. Furthermore the impressive value $f^*_{E-1} = 1.00$ achieved suggests that despite the increased non-convexity with $N$, the required number of PCVTs needed to navigate the energy landscape, in a similar fashion as the random sampling made with gradient based methods, has decreased to 9 (i.e less than one full run).

Figure 5.2. Optimal PCVTs from Example 1: (a) ground state with $E^{min} - 1 \approx 1e-14$ reached by our hybrid method. (b) configuration carrying $E^{ref} - 1 = 0.00287$ achieved by L-BFGS(7). (c) and (d) are the measures of a hybrid run that reached $E^{min}$ amongst the larger batch of 100 runs with $K = 6000$ (the dotted black line designates $E^{ref} - 1$ again).

(a) PCVT achieving $E^{min}$
(b) PCVT with $E^{ref}$
(c) hybrid run reaching $E^{min}$
(d) $R$, $H$ profiles of (c)
Table 1

| K  | N  | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 |
|----|----|-----|-----|-----|-----|-----|-----|
| 8000 | 2029 | 16 | 16 | 16 | 16 | 16 | 16 |

Statistics of Example 1 for N = 973 with a 100 runs of our hybrid method using K = 6000, the values \( E^{\text{max}} - 1, R^\text{max}, H^\text{max} \) as well as \( E^{\text{ref}} - 1, R^\text{ref}, H^\text{ref} \) are bold faced.

Table 2

| K  | N  | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 |
|----|----|-----|-----|-----|-----|-----|-----|
| 8000 | 1000 | 16 | 16 | 16 | 16 | 16 | 16 |

Statistics of Example 2 for N = 2029 with a 100 runs of our hybrid method using K = 8000.

Square torus \( \Omega \). We now work with \( \Omega \) being the square torus and \( N = \{ n \times 1000 \}_{n=1}^{16} \) to investigate the behaviour of our method when the shape effects are tightened.
**Example 3.** For $N = 1000$ with $K = 6000$ (as in Example 1) we implemented a large batch of 1,000 runs of our algorithm so as to have more robust statistics that are presented in Figure 5.5 and in Table 3. We further wish to emphasize, through the monotone skewness in $q$ of the ECDFs and histograms, the increasing performance tendency from one stage to another in this scenario where the contraction and relaxation phases are more constrained by the domain’s shape.

The behaviour is illustrated in the mosaic of PVDs from Figure 5.4 where snapshots of the run achieving $E_{\text{min}} - 1 = 5.27e^{-4}$ were taken. In particular one appreciates how stage after stage:

i) the $\text{MACN-c}$ dynamics make defective (non-hexagonally regular) regions of the PVTs “communicate” with each other by creating a flow between regions with high average of individual energy $E_i$ and ones with sensibly lower value, exhibiting then a clear non-local behaviour.

ii) Lloyd’s algorithm contracts the system while preserving the localization of the defects.

iii) $\text{MACN-δ}$ preserves regions of hexagonal regularity better and better as the energy of the perturbed PCVT diminishes whilst, similarly to $\text{MACN-c}$, creating a non-local “communication” between defects.

At last, we point out that the progressive constriction of the defect “interfaces” we observe after the Lloyd block (the two middle columns on the mosaic) is recurrent across all set ups that were tested, this is simply a consequence of the remarkable navigation of the energy that our hybrid method performs. The video animating the iterations of this particular run can be found in the accompanying supplementary material, with it will appear in more detail how the combination of these symbiotic blocks seem to have a similar effect to a grain boundary evolution algorithm in polycrystals when looking at the produced sequence of PCVTs.

**Example 4.** For $N = 2000$ with $K = 8000$ (as in Example 2) our results for 100 runs are summarized in Table 4. Here P-L-BFGS(20,20) achieved $E_{\text{ref}} - 1 = 0.00495$ which yields the ratio $\tau^{-1} \approx 2.6$. We note that despite the non-negligible performance decrease of $\tau^{-1}$ when compared to Example 2; the values of $f_{E-1}$ above 90% for $X_{q \geq 4}$ indicate that, statistically, the overall comparative efficiency is remarkably similar to what we obtained in the hexagonal torus domain. This suggests that the aforementioned change in $\tau$ could be mainly attributed to the increased rigidity of the square torus.

On the other hand when restricting our attention to the distributions of $E$ gotten thus far from 100 runs or more, we make the point that there is no such discrepancy since our energy scaling shows remarkable robustness for each method (e.g $(E - 1)$ seem to remain comparable regardless of $N$ and $\Omega$, this will be seen as well in the remaining cases).

**Example 5.** When running our algorithm on a total of 100 initial configurations with $K = 8000$ for $N = 3000$ and comparing against P-L-BFGS(20,20), we get $\tau^{-1} \approx 3.1$ as well as values of $f_{E-1}^*$ above 90% for $X_{q \geq 1}$; see Table 5. The increasing relative performance with $N$ is once again apparent.

**Example 6.** At last we consider $N = 4000$ and 100 runs with $K = 12000$, see Table 6. The reader can appreciate how this set up comes to further corroborate the assertions made thus far about the nature of our hybrid algorithm, namely:
i) the monotone decrease of $\langle E - 1 \rangle$ in terms of $q$ for batches of $100$ runs or more. Additionally we’ve seen the robustness of $E$ in terms of $N$ for the PCVTs produced by each method individually, which suggests for example that our hybrid method is able to get on average twice as close to the non-achievable regular hexagonal configuration than the Quasi-Newton methods within $Q = 10$ stages.

ii) as the non-convexity of the problem increases with $N$ it is of course harder to get closer to the ground state (the smallest energies recorded increase regardless of the method), yet $f_{E_{-1}}^*$ surpasses 90% at earlier stages $q$ the larger the number of generators. This shows that the manner our method probes the energy landscape manages to overcome this stiffness remarkably better than gradient based methods combined with the random-like sampling.

Graphics providing visual insight on the results of Examples 2, 4, 5 and 6 can be found in the accompanying supplementary material.
Figure 5.4. Configurations with $N = 1000$ from Example 3: (a) optimal PCVT obtained with Lloyd’s with $E - 1 = 0.00466$, (b) optimal PCVT obtained with L-BFGS(7) having value $E - 1 = 0.00349$, (c) minimal energy configuration obtained amongst the three comparative methods –achieved by P-L-BFGS(20,20)– with $E_{\text{ref}} - 1 = 0.00289$.

The mosaic shows the last iteration of each of the three blocks of our hybrid method across the 10 stages of the run that achieved $E_{\text{min}} - 1 = 5.27 \times 10^{-4}$, the PCVT carrying that minimal value is framed in a red box. These images are to be read starting on the left side of the vertical double black bar and then on the right, each row is for a stage $q = 0,...,9$. Both left columns are the last iteration of the MACN-c blocks, the middle columns contain the PCVTs from the Lloyd blocks and the right columns are their respective MACN-δ displacements. Finally (d) and (e) are the regularity measures’ profiles of the run depicted in the mosaic, i.e the one that achieved $E_{\text{min}}$ amongst the 1000 runs using $K = 6000$. 

| MACN-c | Lloyd block | MACN-δ | MACN-c | Lloyd block | MACN-δ |
|--------|-------------|--------|--------|-------------|--------|
| ![Image](a) Lloyd | ![Image](b) L-BFGS(7) | ![Image](c) P-L-BFGS(20,20) | ![Image](d) hybrid run achieving $E_{\text{min}}$ | ![Image](e) $R_e, H$ profiles associated with (d) |
Figure 5.5. Distributions from Example 3 for $N = 1000$ with $K = 6000$: ECDFs and histograms of regularity measures for the 1,000 hybrid runs as well as for the 100,000 runs of P-L-BFGS(20,20). Further detail on these measures are given in Table 3.
Statistics from Example 3 with $N = 1000$ and $K = 6000$: a thousand runs of our hybrid method versus the comparative algorithms, the values $E_{\min}^m - 1$, $R_{\min}^m$, $H_{\min}$ as well as $E_{\ref}^m - 1$, $R_{\ref}^m$, $H_{\ref}^m$ are bold faced.

| Algorithm | $E_{\min}^m - 1$ | $R_{\min}^m$ | $H_{\min}$ | $E_{\ref}^m - 1$ | $R_{\ref}^m$ | $H_{\ref}^m$ |
|-----------|-----------------|--------------|------------|-----------------|--------------|------------|
| Hybrid    | 0.0057          | 0.0057       | 0.0057     | 0.0057          | 0.0057       | 0.0057     |
| LBFGS     | 0.0057          | 0.0057       | 0.0057     | 0.0057          | 0.0057       | 0.0057     |
| PLBFGS    | 0.0058          | 0.0058       | 0.0058     | 0.0058          | 0.0058       | 0.0058     |

Statistics out of 100 hybrid method runs from Example 4 with $N = 2000$ and $K = 8000$.

| Algorithm | $E_{\min}^m - 1$ | $R_{\min}^m$ | $H_{\min}$ | $E_{\ref}^m - 1$ | $R_{\ref}^m$ | $H_{\ref}^m$ |
|-----------|-----------------|--------------|------------|-----------------|--------------|------------|
| Hybrid    | 0.0057          | 0.0057       | 0.0057     | 0.0057          | 0.0057       | 0.0057     |
| LBFGS     | 0.0057          | 0.0057       | 0.0057     | 0.0057          | 0.0057       | 0.0057     |
| PLBFGS    | 0.0058          | 0.0058       | 0.0058     | 0.0058          | 0.0058       | 0.0058     |

Statistics out of 100 hybrid method runs from Example 5 with $N = 3000$ and $K = 8000$.

| Algorithm | $E_{\min}^m - 1$ | $R_{\min}^m$ | $H_{\min}$ | $E_{\ref}^m - 1$ | $R_{\ref}^m$ | $H_{\ref}^m$ |
|-----------|-----------------|--------------|------------|-----------------|--------------|------------|
| Hybrid    | 0.0057          | 0.0057       | 0.0057     | 0.0057          | 0.0057       | 0.0057     |
| LBFGS     | 0.0057          | 0.0057       | 0.0057     | 0.0057          | 0.0057       | 0.0057     |
| PLBFGS    | 0.0058          | 0.0058       | 0.0058     | 0.0058          | 0.0058       | 0.0058     |

Statistics out of 100 hybrid method runs from Example 6 with $N = 4000$ and $K = 12000$.

| Algorithm | $E_{\min}^m - 1$ | $R_{\min}^m$ | $H_{\min}$ | $E_{\ref}^m - 1$ | $R_{\ref}^m$ | $H_{\ref}^m$ |
|-----------|-----------------|--------------|------------|-----------------|--------------|------------|
| Hybrid    | 0.0057          | 0.0057       | 0.0057     | 0.0057          | 0.0057       | 0.0057     |
| LBFGS     | 0.0057          | 0.0057       | 0.0057     | 0.0057          | 0.0057       | 0.0057     |
| PLBFGS    | 0.0058          | 0.0058       | 0.0058     | 0.0058          | 0.0058       | 0.0058     |

Scope on $R_f$. We have taken $\epsilon$ to be fixed at 0.5% because this value is bounded above by the deviation from $r_{hec}$ gotten from the $\delta$-perturbation of a single generator in the honeycomb PCVT, yet it remains big enough so that the data clearly shows that:

- a higher variation $|\Delta E|$ between consecutive iterations in our method re-
results in a higher $|\Delta R_e|$ than $|\Delta H|$, and this regardless of the block MACN-c/Lloyd/MACN-δ

- we have systematically that $|f^*_E - f^*_R_e| < |f^*_E - f^*_H|$
- the ECDFs of $H$ present larger discontinuity jumps that the ones of $R_e$; meaning that for given $X^*$ the number of computed PCVTs states sharing the value $H(X^*)$ is higher than the one sharing $R_e(X^*)$.

These observations combined point out that $R_e$ is indeed a measure more faithful to $E$ and a better indicator of “well distributed” PCVTs than $H$ is. We provide further insight on this matter in Table 7 through the correlation ratio

$$\varrho := \frac{\sigma_{R_e \text{cov}}(E - 1, H)}{\sigma_{H \text{cov}}(E - 1, R_e)}$$

and in Figure 5.6 through scatter plots of the data from Example 3.

| Es | X_0 | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 | X_7 | X_8 | X_9 |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1  | 1.0621 | 1.0924 | 1.0838 | 1.9875 | 1.0550 | 1.0687 | 1.0674 | 1.0501 | 1.0523 | 1.0578 |
| 2  | 1.0681 | 1.0934 | 1.0742 | 1.1228 | 1.1250 | 1.1431 | 1.1294 | 1.1924 | 1.1315 | 1.1033 |
| 3  | 1.0686 | 1.0714 | 1.0746 | 1.0744 | 1.0667 | 1.0689 | 1.0964 | 1.0955 | 1.0416 | 1.0741 |
| 4  | 1.0995 | 1.0690 | 1.0646 | 1.0964 | 1.0759 | 1.0649 | 1.0743 | 1.0806 | 1.0960 | 1.0849 |
| 5  | 1.0644 | 1.0608 | 1.0752 | 1.0898 | 1.0882 | 1.0328 | 1.0339 | 1.0429 | 1.0435 | 1.0430 |
| 6  | 1.0966 | 1.0546 | 1.0742 | 1.0803 | 1.0724 | 1.0726 | 1.0602 | 1.0447 | 1.0508 | 1.0449 |

Table 7

Correlation ratios $\varrho$ for each method from the PCVT data presented in Examples 1 through 6

| Lloyd | LIFPGS | PLIFPGS |
|-------|--------|---------|
| 1.0710 | 1.0663 | 1.0666 |
| 1.0417 | 1.0688 | 1.0683 |
| 1.0682 | 1.0674 | 1.0678 |
| 1.0869 | 1.0692 | 1.0673 |
| 1.0820 | 1.0689 | 1.0679 |
| 1.0842 | 1.0688 | 1.0665 |

Figure 5.6. Scatter plots of $E$ vs. $H$ and $E$ vs. $R_e$ displaying the same data as Figures 5.5 (a) through (f) from Example 3.

6. Closing Remarks and future directions. The contributions of this paper were two-fold: first the introduction of the isoperimetric ratio via $R_e$ as an indicator of low energy CVTs and second, most importantly, to our knowledge, the first coupling algorithm capable of successfully probing the CVT energy landscape in a completely deterministic way while keeping $N$ fixed.

We point out that while we prioritized simplicity of the method in this paper, the algorithm’s performance could be further improved, if needed be, by: i) using initial quasi-random distributions, ii) replacing Lloyd’s with other gradient based descent
methods that satisfy Wolfe conditions, and most importantly iii) by introducing a suitable decay in the sequence \(\{K_q\}_{q=0}^{Q-1}\) (possibly adapted to \(\{E(X_q^*)\}_{q=0}^{Q-1}\)). The point made is that even with the crude tunings made on \(K_q \equiv K\) in §5, the resulting regularity measures are remarkable.

Finally, in terms of our direct CVT/optimal quantization application, it would be natural to explore our global method:

- on the 3D square torus wherein one would expect the appearance of the BCC lattice and truncated octahedron Voronoi cells;
- on the 2-sphere where one would expect the lowest energy state to be the “soccer ball” structure consisting of \(N-12\) regular hexagons and 12 regular pentagons.

A different related question is the inclusion of an underlying inhomogeneous probability densities \(\rho\) over \(\Omega\) wherein the energy (1.2) takes the form

\[
F(X) = \sum_{i=1}^{N} \int_{V_i} ||y - x_i||^2 \rho(y) \, dy.
\]

However, here it is unclear how to choose the distance \(\delta\) in the MACN annealing step.

REFERENCES

[1] E. Barnes and N. Sloane, *The optimal lattice quantizer in three dimensions*, SIAM Journal on Algebraic Discrete Methods, 4 (1983), pp. 30–41.
[2] E. Bormashenko, M. Frenkel, and I. Legchenkova, *Is the voronoi entropy a true entropy? comments on "entropy, shannon’s measure of information and boltzmann’s h-theorem*, Entropy 2017, 19, 48, Entropy, 21 (2019), p. 251.
[3] E. Bormashenko, M. Frenkel, A. Vilk, I. Legchenkova, A. A. Fedorets, N. E. Aktaev, L. A. Dombrovsky, and M. Nosonovsky, *Characterization of self-assembled 2d patterns with voronoi entropy*, Entropy, 20 (2018), p. 956.
[4] M. Caroli and M. Teillaud, *Computing 3d periodic triangulations*, in European Symposium on Algorithms, Springer, 2009, pp. 59–70.
[5] R. Choksi and X. Y. Lu, *Bounds on the geometric complexity of optimal centroidal voronoi tessellations in 3d*, Comm. Math Phys, (to appear).
[6] P. N. Chung, M. Fernandez, N. Shah, L. S. Vieira, E. Wikner, et al., *Perimeter-minimizing pentagonal tilings*, Involv: A Journal of Mathematics, 7 (2014), pp. 453–478.
[7] J. Conway and N. Sloane, *Voronoi regions of lattices, second moments of polytopes, and quantization*, IEEE Transactions on Information Theory, 28 (1982), pp. 211–226.
[8] Q. Du and M. Emelianenko, *Acceleration schemes for computing centroidal voronoi tessellations*, Numerical linear algebra with applications, 13 (2006), pp. 173–192.
[9] Q. Du, M. Emelianenko, and L. Ju, *Convergence of the lloyd algorithm for computing centroidal voronoi tessellations*, SIAM journal on numerical analysis, 44 (2006), pp. 102–119.
[10] Q. Du, V. Faber, and M. Gunzburger, *Centroidal voronoi tessellations: Applications and algorithms*, SIAM review, 41 (1999), pp. 637–676.
[11] Q. Du, M. Gunzburger, and L. Ju, *Advances in studies and applications of centroidal voronoi tessellations*, Numerical Mathematics: Theory, Methods and Applications, 3 (2010), pp. 119–142.
[12] Q. Du and D. Wang, *The optimal centroidal voronoi tessellations and the gershho’s conjecture in the three-dimensional space*, Computers & Mathematics with Applications, 49 (2005), pp. 1355–1373.
[13] M. Emelianenko, L. Ju, and A. Rand, *Nondegeneracy and weak global convergence of the lloyd algorithm in r^n*, SIAM Journal on Numerical Analysis, 46 (2008), pp. 1423–1441.
[14] A. Gersho, *Asymptotically optimal block quantization*, IEEE Transactions on information theory, 25 (1979), pp. 373–380.
[15] S. Graf and H. Luschgy, Foundations of quantization for probability distributions, Springer, 2007.
[16] Gruber, A short analytic proof of fejes tóth’s theorem on sums of moments, aequationes mathematicae, 58 (1999), pp. 291–295.
[17] W. W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods, Pacific Journal of Optimization, 2 (2006), pp. 35–58.
[18] J. H. Halton, On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, Numerische Mathematik, 2 (1960), pp. 84–90.
[19] J. M. Hammersley, Monte carlo methods for solving multivariable problems, Annals of the New York Academy of Sciences, 86 (1960), pp. 844–874.
[20] J. C. Hateley, H. Wei, and L. Chen, Fast methods for computing centroidal voronoi tessellations, Journal of Scientific Computing, 63 (2015), pp. 185–212.
[21] M. Iri, K. Murota, and T. Ohya, A fast voronoi-diagram algorithm with applications to geographical optimization problems, in System modelling and optimization, Springer, 1984, pp. 273–288.
[22] L. Jiang, R. H. Byrd, E. Eskow, and R. B. Schnabel, A preconditioned l-bfgs algorithm with application to molecular energy minimization, tech. rep., COLORADO UNIV AT BOULDER DEPT OF COMPUTER SCIENCE, 2004.
[23] L. Ju, Q. Du, and M. Gunzburger, Probabilistic methods for centroidal voronoi tessellations and their parallel implementations, Parallel Computing, 28 (2002), pp. 1477–1500.
[24] C.-J. Lin and J. J. More, Incomplete cholesky factorizations with limited memory, SIAM Journal on Scientific Computing, 21 (1999), pp. 24–45.
[25] Y. Liu, W. Wang, B. Lévy, F. Sun, D.-M. Yan, L. Lu, and C. Yang, On centroidal voronoi tessellation-energy smoothness and fast computation, ACM Transactions on Graphics (ToG), 28 (2009), p. 101.
[26] S. Lloyd, Least squares quantization in pcm, IEEE transactions on information theory, 28 (1982), pp. 129–137.
[27] L. Lu, F. Sun, H. Pan, and W. Wang, Global optimization of centroidal voronoi tessellation with monte carlo approach, IEEE transactions on visualization and computer graphics, 18 (2012), pp. 1880–1890.
[28] T. McSweeney, Modified cholesky decomposition and applications, Master’s thesis, University of Manchester, 2017.
[29] Y. Men, Z. Shen, D. Khan, and D.-M. Yan, Improving regularity of the centroidal voronoi tessellation, in ACM SIGGRAPH 2018 Posters, ACM, 2018, p. 66.
[30] M. Moriguchi and K. Sugihara, A new initialization method for constructing centroidal voronoi tessellations on surface meshes, in 2006 3rd International Symposium on Voronoi Diagrams in Science and Engineering, IEEE, 2006, pp. 159–165.
[31] D. Newman, The hexagon theorem, IEEE Transactions on information theory, 28 (1982), pp. 137–139.
[32] H. Niederreiter, Random number generation and quasi-Monte Carlo methods, vol. 63, Siam, 1992.
[33] J. Nocedal and S. Wright, Numerical optimization, Springer Science & Business Media, 2nd ed., 2006.
[34] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, Spatial tessellations: concepts and applications of Voronoi diagrams, vol. 501, John Wiley & Sons, 2009.
[35] J. Quinn, F. Sun, F. C. Langbein, Y.-K. Lai, W. Wang, and R. R. Martin, Improved inititalization for centroidal voronoi tessellation and optimal delaunay triangulation, Computer-Aided Design, 44 (2012), pp. 1062–1071.
[36] The CGAL Project, CGAL User and Reference Manual, CGAL Editorial Board, 5.0 ed., 2019.
[37] L. F. Tóth, Lagerungen in der Ebene auf der Kugel und im Raum, Springer-Verlag, 1953.
[38] L. Wang, F. Hétrony-Wheeler, and E. Boyer, A hierarchical approach for regular centroidal voronoi tessellations, in Computer Graphics Forum, vol. 35, Wiley Online Library, 2016, pp. 152–165.
[39] D.-M. Yan, K. Wang, B. Lévy, and L. Alonso, Computing 2d periodic centroidal voronoi tessellation, in 2011 Eighth International Symposium on Voronoi Diagrams in Science and Engineering, IEEE, 2011, pp. 177–184.
[40] J. Zhang, M. Emelianenko, and Q. Du, Periodic centroidal voronoi tessellations., International Journal of Numerical Analysis & Modeling, 9 (2012).
Appendix A. Scope on MACN-$\delta$. As defined in (1.3), the value of $\delta$ is crucial since it has the peculiarity of making our perturbation stage preserve a certain regularity in the structure of the tessellation, however the direction of the perturbation seems to be of primary importance compared to the step size when the latter is fixed. To gain insight on this matter we define three variants of our perturbation step:

1. Inspired by the relation between $\delta$ and the intrinsic length-scale of the regular hexagonal lattice, the first variant consists of moving away from the closest neighbor by the length-scale proper to each cell in the PCVT, i.e the perturbation follows

\[
x_i \leftarrow x_i + \frac{|V_i|}{|\partial V_i|} \frac{x_i - x_j^*}{||x_i - x_j^*||}, \quad i = 1, ..., N
\]

2. The next variant contemplates $\delta$ as in (1.3) but choses a random neighbor $x_j, j \in N_i$ to move away from, thus not necessarily being the closest one.

3. Finally, we consider each generator moving by the distance $\delta$ and at a random angle $\theta_i \in [0, 2\pi)$ taken from the uniform distribution.

Figure A.1 illustrates how close the performance of our variants of the MACN-$\delta$ step are from one another but with particular distinction of the random angle $\theta_i$ perturbation, being then of some reassurance that our original guided dislocation of PCVTs is better suited than some random search in a $\delta$–vicinity.

![Comparison in the energy performance of the original MACN-$\delta$ step and its three variants across stages of our hybrid method with $N = 1000$ and $K = 6000$. The joint markers represent averages while the isolated ones represent minima over 100 runs.](image)
Appendix B. Explicit pseudo code for our hybrid method.

Algorithm B.1 Hybrid algorithm over $Q$ stages with Lloyd’s subroutine

**Input:** 1) initial set of generators $X = \{x_i\}_{i=1}^N$; 2) number of stages $Q$; 3) integer sequence $\{K_q\}_{q=0}^{Q-1}$ for the $MACN-c$ blocks and 4) $tol$ for convergence to a PCVT

for $q = 0 : Q - 1$ do

I. $MACN-c$:
   for $k = 0 : K_q - 1$ do
      compute PDT($X$) and extract $\{\mathcal{N}_i\}_{i=1}^N$
      compute PVT($X$) as the dual of PDT($X$)
      for $i = 1 : N$ do
         find one index $j^*_i$ solving $\min_{j \in \mathcal{N}_i} ||x_i - x_j||$
         compute the centroid $c_i$
      end for
      update $X$: $x_i \leftarrow x_i \forall i$ with (4.1)
   end for

II. Reaching criticality:
   $[X^*_q, \text{PDT}(X^*_q), \text{PCVT}(X^*_q)]$=Lloyd($X$, $tol$)

III. $MACN-\delta$:
   if $q < Q - 1$ then
      from PDT($X^*_q$) extract $\{\mathcal{N}_i\}_{i=1}^N$
      for $i = 1 : N$ do
         find one index $j^*_i$ solving $\min_{j \in \mathcal{N}_i} ||x_i - x_j||$
      end for
      get new $X$ by $\delta$-perturbing $X^*_q$: $x_i \leftarrow x_i \forall i$ with (4.2)
   end if

end for

**Output:** $\{X^*_q\}_{q=0}^{Q-1}$, a collection of stable local minimizers of $E$ and their corresponding PCVTs and PDTs.

**subroutine** $[X, \text{PDT}(X), V(X)]$=Lloyd($X$, $tol$)

set $\text{diff} = \text{Inf}$
while $\text{diff} > tol$ do
   compute $V(X)$ (i.e PVT($X$))
   for $i = 1 : N$ do
      compute the centroid $c_i$ and area $|V_i|$
   end for
   update $X$: $x_i \leftarrow c_i \forall i$
   $\text{diff} = ||D E||/N$
end while
end subroutine
Appendix C. Explicit pseudo code for P-L-BFGS method.

Algorithm C.1 P-L-BFGS($M, T$)

Prior definitions: to ease notation we define at each iteration $k$

\[
s_k := X^{(k+1)} - X^{(k)}; \quad y_k := DE(X^{(k+1)}) - DE(X^{(k)})
\]

\[
\rho_k := \frac{1}{y_k s_k} \quad \text{and} \quad H_0^{(k)} := \frac{s_{k-1}^T y_{k-1}}{y_{k-1} y_{k-1}} I
\]

Input: i) initial iterate $X^{(0)}$; ii) integer parameters $M$ and $T$; iii) tolerance $tol$ for convergence

set $k = 0$
set $\text{diff} = \text{Inf}$
while $\text{diff} > tol$ do
set $q = DE(X^{(k)})$

1st L-BFGS update
for $i = k - 1 : -1 : k - M$ do
\[
a_i = \rho_i s_i^T q
\]
$q \leftarrow q - a_i y_i$
end for

Redirect search direction
if $k \mod T = 0$ then
construct preconditioner matrix $\tilde{A}_k$ and solve the system $\tilde{A}_k r = q$
else
construct $H_0^{(k)}$ and set $r = H_0^{(k)} q$
end if

2nd L-BFGS update
for $i = k - M : k - 1$ do
$r \leftarrow r + s_i (a_i - \rho_i y_i^T r)$
end for

set descent direction $p^{(k)} = -r$
update iterate $X^{(k+1)} = X^{(k)} + a^{(k)} p^{(k)}$ where $a^{(k)}$ is a step length satisfying the strong Wolfe conditions

if $k > M$ then
erase the tuple $\{s_{k-M}, y_{k-M}\}$
compute and store $\{s_k, y_k\}$
end if

\[
\text{diff} = \frac{||DE||}{N}
\]
k $\leftarrow$ $k + 1$
end while

Output: $X^*$, a stable local minimizer of $E$ and its corresponding PCVT and PDT