

# Efficient path generation with reduced coordinates

Renjie Chen · Craig Gotsman · Kai Hormann

---

## Abstract

Path generation is an important problem in many fields, especially robotics. One way to create a path between a source point  $z$  and a target point  $y$  inside a complex planar domain  $\Omega$  is to define a non-negative distance function  $d(y, z)$ , such that following the negative gradient of  $d$  (by  $z$ ) traces out such a path. This presents two challenges: (1) The mathematical challenge of defining  $d$ , such that  $d(y, z)$  has a single minimum at  $z = y$  for any fixed  $y$ , because the gradient-descent path may otherwise terminate at a local minimum before reaching  $y$ ; (2) The computational challenge of defining  $d$ , such that it can be computed efficiently. Using the concepts of *harmonic measure* and *f-divergence*, we show how to assign a set of *reduced coordinates* to each point in  $\Omega$  and to define a family of distance functions based on these coordinates, such that both the mathematical and the computational challenge are met. Since in practice, especially in robotics applications, the path is often restricted to follow the edges of a discrete network defined on a finite set of *sites* sampled from  $\Omega$ , any method that works well in the continuous setting must be discretized appropriately to preserve the important properties of the continuous case. We show how to define a network connecting a finite set of sites, such that a *greedy routing* algorithm, which is the discrete equivalent of continuous gradient descent, based on our reduced coordinates is guaranteed to generate a path in the network between any two sites. In many cases, this network is close to a planar graph, especially if the set of sites is dense. Guaranteeing the existence of a greedy route between any two points in the graph is a significant advantage in practical applications, avoiding the complexity of other path-planning methods, such as the shortest-path and A\* algorithms. While the paths generated by our algorithm are not the shortest possible, in practice we found that they are close to that.

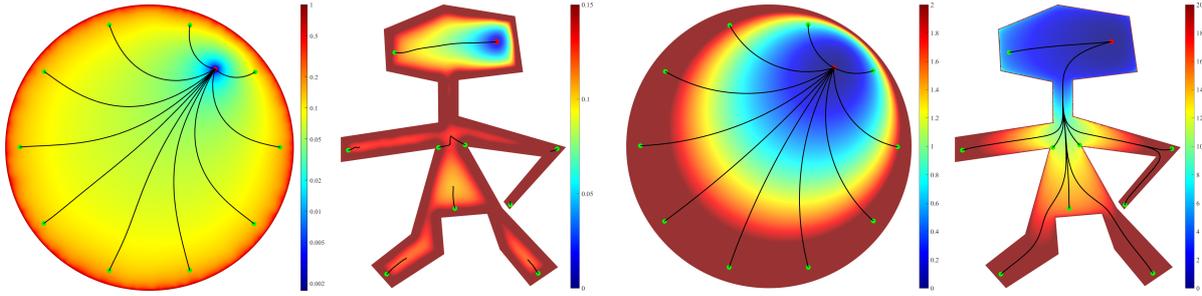
## Citation Info

*Journal*  
Computer Graphics  
Forum  
*Volume*  
37(5), August 2018  
*Pages*  
37–48  
*Note*  
Proceedings of SGP

---

## 1 Introduction

Path generation in complex planar domains is an important and challenging problem in robotic navigation. The objective is for the robot to move from one point (the source) in the domain to another (the target) along a short and smooth path that stays within the domain and avoids obstacles. Practical implementations discretize the domain as a dense graph and “route” along the edges of this graph, ideally generating such paths. Common algorithms are the classical Dijkstra algorithm for shortest-path computation and the more sophisticated A\* [11] and its variants. The interested reader is referred to Souissi et al. [22] for a recent survey of these methods. The simple shortest-path algorithm, while generating an optimal path, can be notoriously slow, and in principle may search the entire graph in order to find the path. Using heuristic functions to provide a lower-bound estimate of the remaining distance in the A\* algorithm may speed this up by avoiding traversal of irrelevant portions of the graph, but, depending on the accuracy of the heuristic function, may still incur significant overhead. In terms of complexity, the best situation to hope for is the existence of a distance function, which can guide the robot from the source to the target by systematically minimizing this distance. Ideally, a path is generated by *gradient descent* on this function, where the robot moves from a node in the graph to the neighbouring node which reduces the distance function the most, until the target is reached. This method, sometimes called *greedy routing*, is an extremely simple algorithm. It requires no priority queue and incurs no extra overhead of exploring other nodes of the graph, as opposed to A\* and its variants. The small price to pay is that the paths are not the shortest possible. But shortest paths may not be desirable anyway at times, since they have sharp corners. However, greedy routing assumes quite optimistically that the distance function has no local minima, otherwise the procedure could get stuck there (namely, reach a vertex from which all neighbours increase the distance to the target). Indeed, if the shortest-path distance from the target to every other node in the graph was known in advance, following a gradient-descent path on this distance would immediately (by definition) trace out the shortest path. As this distance is obviously not known in advance, a second best is to use a different distance function, which is easy to compute, void of local minima and generates paths not too much longer than the shortest path. Defining and implementing such distance functions is the topic of this paper.



**Figure 1:** A smooth path from a source point  $z$  (green) to a target point  $y$  (red) inside a planar domain can be generated by following the negative gradient of a shape-aware distance function  $d(y, z)$  that measures some distance between the Poisson kernels at  $y$  and  $z$ . While the  $L^2$ -distance (left) works well for simple shapes, for complex domains it has local minima where the gradient-descent path terminates. In contrast, the  $f$ -divergence (right) is guaranteed to be void of local minima and can be implemented efficiently using reduced coordinates.

The idea of path-planning based on gradient descent is not new. A well-known family of path generation algorithms, inspired by the physics of electrical force fields, is based on *potential functions*. These were first proposed by Khatib [13] and developed by a number of authors [15, 14, 20, 7] soon after. As described above, the main idea is, given the target point, to construct a scalar function on the domain, such that a path to the target point from any other source point may be obtained by following the negative gradient of the function. For this to work as planned, the scalar function must have a global minimum at the target and be void of local minima elsewhere in the domain. Other critical points, such as saddles, are undesirable but not fatal, since a negative gradient can still be detected by probing around the saddle. Designing and computing potential functions for complex planar domains, possibly with holes (representing obstacles), has been a topic of intense activity for decades. Perhaps the most elegant type of potential function is the harmonic function [1], which has very useful mathematical properties, most notably the guaranteed absence of local minima, in both the continuous and discrete settings. Alas, the main problem preventing widespread use of this potential function is the high complexity of computing the function, essentially the solution of a very large system of linear equations on a discretization of the domain every time the target point is changed, and the fact that very high precision numerical methods are required, as the functions are almost constant, especially in regions distant from the target. A recent paper by Chen et al. [4] addresses the first of these issues. They describe a new family of continuous distance functions, which, while quite distinct from the harmonic potential function, also possess no local minima and generate exactly the same gradient descent paths as the harmonic potential. The present paper expands on these functions, in particular for the discrete setting. Since the discrete version of these distance functions are not completely void of local minima, we show how to remedy that. More importantly, we show how to compute variants of these distance functions, based on the concept of *reduced coordinates*, extremely efficiently.

Chen et al. [4] use the concept of harmonic measure [10] and its conformal invariance to define a family of shape-aware distance functions  $d_f(y, z)$  on a bounded simply-connected planar domain  $\Omega$ . These  *$f$ -divergence distances* (see Section 3) are defined as the  *$f$ -divergence* [8, 16] of the two Poisson kernels of  $\Omega$  at  $y$  and  $z$ , where  $f$  is a strictly convex real function. Figure 1 shows examples of gradient-descent paths for two bounded simply-connected domains, using the  $f$ -divergence (right) and the  $L^2$ -distance (left) between the Poisson kernels. While the latter can have local minima where the paths may terminate, the paths derived from the  $f$ -divergence always reach the target. Chen et al. [4] prove that  $d_f(y, z)$  is symmetric, subharmonic, has a global minimum at  $z = y$  and no local minima, and that the paths are *invariant to  $f$* . This is because only the *magnitude* of  $\nabla_z d_f(y, z)$  depends on  $f$ , *but not its direction*. Furthermore, the  $f$ -divergence distance paths are identical to those generated by following the negative gradient of the harmonic potential function.

In practice, we usually consider a discretization of the domain  $\Omega$ , that is, a triangulation  $T$  with  $k$  vertices, where  $k$  can be on the order of hundreds of thousands, and all computations are done on this triangulation. Given a source point  $z \in T$  and a target point  $y \in T$ , path generation using the harmonic potential function requires solving a large  $k \times k$  system of sparse linear equations that stem from the standard Finite Element Method (FEM) discretization of Laplace's equation. This is usually done by computing the Cholesky decomposition of the system matrix in a preprocessing step and then at query time, given a target vertex  $y$ , using back-substitution to compute the distances from *all*  $x \in T$  to  $y$  at once in  $\mathcal{O}(k \log k)$  time, because the Cholesky factor has  $\mathcal{O}(k \log k)$  non-zero entries. In contrast, for  $f$ -divergence distance paths,

the Poisson kernels are determined at *all* the vertices of  $T$  in a preprocessing step, resulting in  $k$  *coordinate vectors* (one per vertex). These vectors contain one value for each boundary point of  $T$  and typically have length  $\mathcal{O}(\sqrt{k})$ . The path from  $z$  to  $y$  can then be found by computing only the distances from the path vertices and their neighbours to  $y$ , where each distance calculation requires  $\mathcal{O}(\sqrt{k})$  operations [4].

While a cost of  $\mathcal{O}(\sqrt{k})$  per distance calculation does not seem too expensive, in practice it may still be too much for real-time performance. Furthermore, the coordinate vectors generated by the preprocessing procedure must be stored for each of the  $k$  vertices of the triangulation  $T$ , implying an  $\mathcal{O}(k\sqrt{k})$  storage requirement, which could be prohibitive. This paper addresses both issues.

## 2 Contribution

This paper makes two main contributions. The first contribution is a more practical approach to path generation based on  $f$ -divergence distances. The key idea is to replace each of the  $k$  coordinate vectors with a very small set of  $n \ll \mathcal{O}(\sqrt{k})$  *reduced coordinates* (see Section 4), without losing the important property that the distance function has no local minima (see Section 5). This reduces the cost of the distance calculations from  $\mathcal{O}(\sqrt{k})$  to  $\mathcal{O}(n)$ . The small price to pay is that the generated gradient-descent paths may not be as natural as before, and we lose the properties of symmetry of the distance function and independence of the gradient direction on  $f$ .

The second contribution shows how to use reduced  $f$ -divergence distances in a purely discrete setting. While reducing the size of the coordinate vector from  $\mathcal{O}(\sqrt{k})$  to  $n$  makes for an efficient computation of the distance function  $d_f$ , the actual path is generated using the triangulation  $T$ , implying a storage requirement of  $\mathcal{O}(kn)$ . In practice it would often be much more efficient to generate paths on a sparse network of  $m \ll k$  sites in  $\Omega$ . This requires building a suitable network of edges between the sites, one that supports *greedy routing*. In other words, for any site  $x$ , there must always be a neighbour of  $x$  with a smaller distance to the target  $y$ , which is the discrete analog of the continuous property of not having local minima. In Section 7 we describe an algorithm for building such a graph.

To illustrate the advantage of both contributions, consider a typical triangulation  $T$  of  $\Omega$  with  $k = 200,000$  vertices and 700 boundary vertices. In this case, the algorithm of Chen et al. [4] requires storing  $1.4 \times 10^8$  real values, and each distance calculation calls for 700 basic calculations. In contrast, using  $n = 30$  reduced coordinates and a network with  $m = 300$  sites requires storing only 9,000 real values and each distance calculation comes at the cost of just 30 basic calculations.

## 3 The $f$ -divergence distance

The  $f$ -divergence distance proposed in [4] is based on the concept of  $f$ -divergence, which was first introduced by Kullback and Leibler [16] and later generalized by Csiszár [8], for measuring the difference between two probability distributions.

**Definition 1.** Let  $f$  be a strictly convex function such that  $f(1) = 0$  and  $p, q: E \rightarrow [0, 1]$  be two real functions on some domain  $E$  such that  $\int_E p(t) dt = \int_E q(t) dt = 1$ . The  $f$ -divergence from  $q$  to  $p$  is

$$d_f(p, q) = \int_E p(t) f\left(\frac{q(t)}{p(t)}\right) dt.$$

It is well known that  $d_f(p, q) \geq 0$  and

$$d_f(p, q) = 0 \quad \iff \quad p = q,$$

but  $d_f$  is not necessarily a metric. Many instances of  $f$  have been proposed over the years, each suitable for some specific application, mostly in probability theory, statistics, and information theory. The interested reader is referred to [17] for a survey of the possibilities.

In this context, it is noteworthy to recall the concept of the dual function

$$f^*(x) = x f(1/x).$$

For example, if  $f(x) = -\log x$ , then  $f^*(x) = x \log x$ , and if  $f(x) = |1 - x|$ , then  $f^*(x) = f(x)$ . It is well known that

1.  $f$  is (strictly) convex if and only if  $f^*$  is (strictly) convex;
2.  $d_f(p, q) = d_{f^*}(q, p)$ ;
3.  $f(1) \leq d_f(p, q) \leq f(0) + f^*(0)$ .

The  $f$ -divergence *distance* further relies on the *Poisson kernel* of  $\Omega$  at  $y \in \Omega$ ,

$$P(t, y) = \frac{\partial D}{\partial n}(t, y), \quad t \in \partial\Omega,$$

which is the derivative of the Green's function  $D(x, y)$  for the Laplace operator with Dirichlet boundary conditions at the boundary point  $t$ , in the direction of the unit outer normal  $n$  at  $t$  [10]. The  $f$ -divergence distance is then defined as the  $f$ -divergence of two Poisson kernels.

**Definition 2.** Let  $f$  be a strictly convex function such that  $f(1) = 0$  and  $y, z \in \Omega$ . The  $f$ -divergence distance from  $z$  to  $y$  is

$$d_f(y, z) = \oint_{\partial\Omega} P(t, y) f\left(\frac{P(t, z)}{P(t, y)}\right) dt. \quad (1)$$

Although the  $f$ -divergence of two probability functions is not necessarily symmetric, the special nature of the  $f$ -divergence distance implies that it is symmetric. In fact,  $f$  and  $f^*$  generate *identical* divergence distances, hence

$$d_f(y, z) = d_{f^*}(y, z) = d_f(z, y).$$

While the  $f$ -divergence distance is symmetric, it is not a metric in general, since it may fail to satisfy the triangle inequality.

For path-generation purposes, the gradient of the  $f$ -divergence distance plays a key role. Chen et al. [4] show that for bounded and simply-connected domains the gradient field vanishes only at the target,

$$\nabla_z d_f(y, z) = 0 \quad \iff \quad z = y, \quad (2)$$

and that its direction is independent of  $f$ , that is,

$$\arg(\nabla_z d_f(y, z)) = \arg(\nabla_z d_g(y, z)),$$

for any two strictly convex functions  $f$  and  $g$ .

## 4 Reduced coordinates

Since the Poisson kernel satisfies

$$\oint_{\partial\Omega} P(t, z) dt = 1$$

and

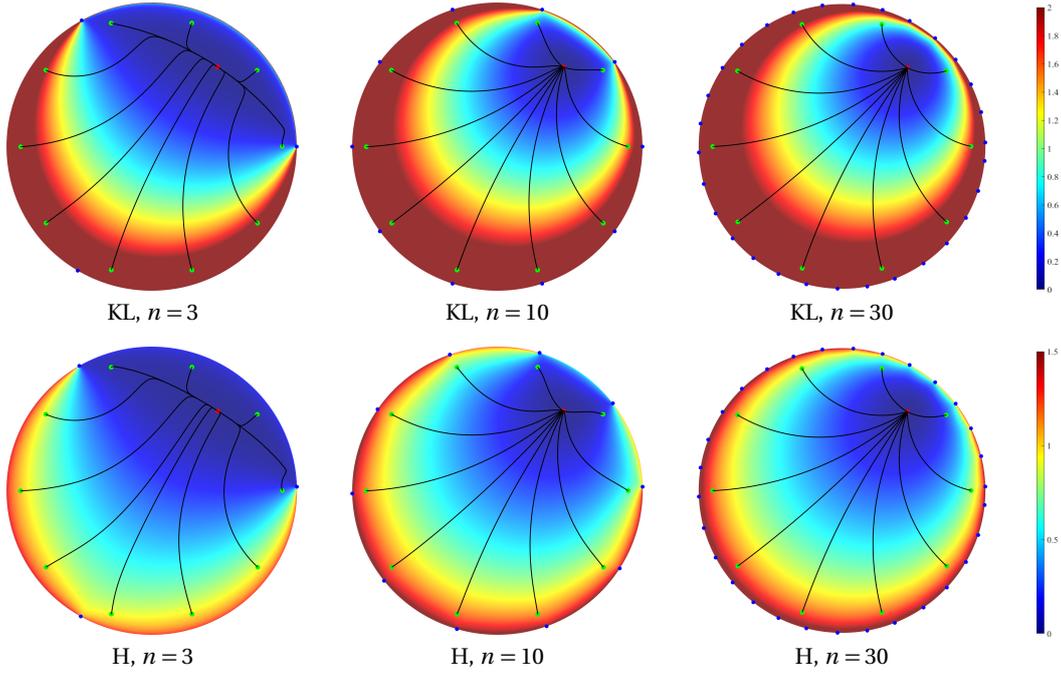
$$\oint_{\partial\Omega} P(t, z) t dt = z,$$

that is, it reproduces constant and linear functions, it can be viewed as a continuous *coordinate vector* for  $z \in \Omega$ , which can be discretized as follows. Let  $t_1, \dots, t_n \in \partial\Omega$  be a sequence of  $n$  points along the domain boundary. We then partition  $\partial\Omega$  into the  $n$  segments  $E_j = [t_j, t_{j+1}]$ ,  $j = 1, \dots, n$ , where  $t_{n+1}$  is identified with  $t_1$  and define the *reduced coordinates*  $(\phi_1(z), \dots, \phi_n(z))$  of  $z \in \Omega$  as

$$\phi_j(z) = \int_{E_j} P(t, z) dt, \quad j = 1, \dots, n. \quad (3)$$

The quantity  $\phi_j(z)$  is also called the *harmonic measure* of  $E_j$  at  $z$  [10]. Using reduced coordinates, we can now discretize the  $f$ -divergence distance in (1), and define the *reduced  $f$ -divergence distance* from  $z$  to  $y$  as

$$d_f(y, z) = \sum_{j=1}^n \phi_j(y) f\left(\frac{\phi_j(z)}{\phi_j(y)}\right). \quad (4)$$



**Figure 2:** Paths from several source points (green) to a common target point (red), generated using reduced KL (top) and reduced H (bottom) divergence distance with respect to a partition of the boundary defined by the  $n$  boundary points (blue) for different  $n$ . Note that the KL paths are in general not identical to the H paths, but both converge to the original  $f$ -divergence paths (based on the continuous coordinates) in Figure 1 (right) as  $n$  increases.

## 5 The divergence gradient theorem

Our main theoretical result is that the reduced  $f$ -divergence distance can be used for generating gradient-descent paths.

**Theorem 1** (Divergence Gradient Theorem). *Let  $\Omega$  be a bounded and simply-connected domain,  $f$  be a strictly convex function, and  $n \geq 3$ . Then the reduced  $f$ -divergence distance (4) has property (2).*

We prove Theorem 1 for the special case of the unit disk with the target point at the origin in Appendix A (see Theorem A.1) and now show how to generalize it to arbitrary bounded and simply-connected domains. The key is the conformal invariance of the  $f$ -divergence distance, which in turn is implied by the well-known conformal invariance of the harmonic measure of  $E \subset \partial\Omega$  at  $z$ ,

$$\phi^\Omega(E, z) = \int_E P^\Omega(t, z) dt.$$

**Theorem 2** ([10]). *Let  $\Omega_1$  and  $\Omega_2$  be two bounded and simply-connected domains and  $C: \Omega_1 \rightarrow \Omega_2$  be a conformal map between them such that  $C(\partial\Omega_1) = \partial\Omega_2$ . Then,*

$$\phi^{\Omega_2}(C(E), C(z)) = \phi^{\Omega_1}(E, z).$$

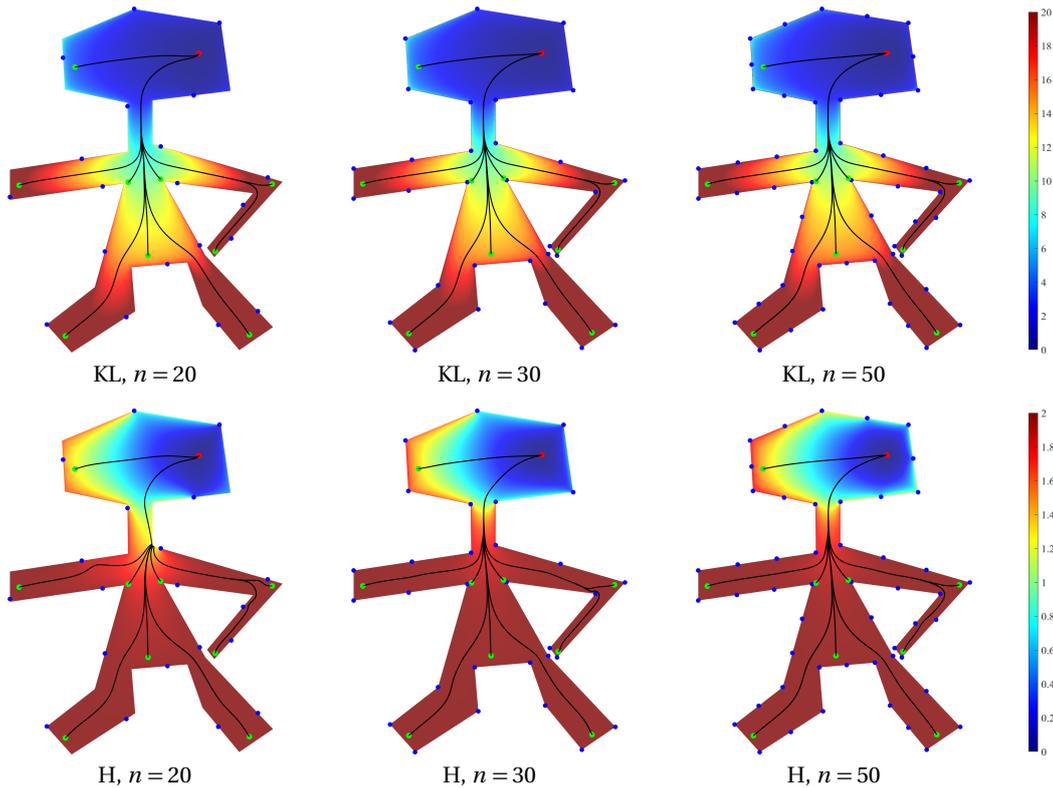
**Theorem 3.** *Let  $C: \Omega_1 \rightarrow \Omega_2$  be a conformal map and  $d_f^{\Omega_1}: \Omega_1 \times \Omega_1 \rightarrow \mathbb{R}$  and  $d_f^{\Omega_2}: \Omega_2 \times \Omega_2 \rightarrow \mathbb{R}$  be the  $f$ -divergence distances for  $\Omega_1$  and  $\Omega_2$ , respectively. Then,*

$$d_f^{\Omega_2}(C(y), C(z)) = d_f^{\Omega_1}(y, z).$$

*Proof.* By the definition of  $d_f$  in (4) and Theorem 2, we have

$$d_f^{\Omega_2}(C(y), C(z)) = \sum_{j=1}^n \phi^{\Omega_2}(C(E_j), C(y)) f\left(\frac{\phi^{\Omega_2}(C(E_j), C(z))}{\phi^{\Omega_2}(C(E_j), C(y))}\right) = \sum_{j=1}^n \phi^{\Omega_1}(E_j, y) f\left(\frac{\phi^{\Omega_1}(E_j, z)}{\phi^{\Omega_1}(E_j, y)}\right) = d_f^{\Omega_1}(y, z)$$

□



**Figure 3:** Same as Figure 2, for a non-convex polygonal domain. Convergence to the original  $f$ -divergence paths in Figure 1 (right) is more rapid for the KL divergence than for the H divergence.

Now we are ready to prove Theorem 1 for any bounded and simply-connected domain by conformally mapping it to the unit disk.

*Proof of Theorem 1.* For a given target point  $y \in \Omega$ , the Riemann Mapping Theorem [10] implies that there exists a conformal map  $C : \Omega \rightarrow D$ , where  $D$  is the unit disk, such that  $\partial\Omega$  is mapped to  $\partial D$  and  $C(y) = 0$ . By Theorem 3,

$$d_f^\Omega(y, z) = d_f^D(0, C(z)).$$

The gradients of the two distance functions with respect to their second argument are related by the chain rule for holomorphic functions. Dropping the first (fixed) argument, we get

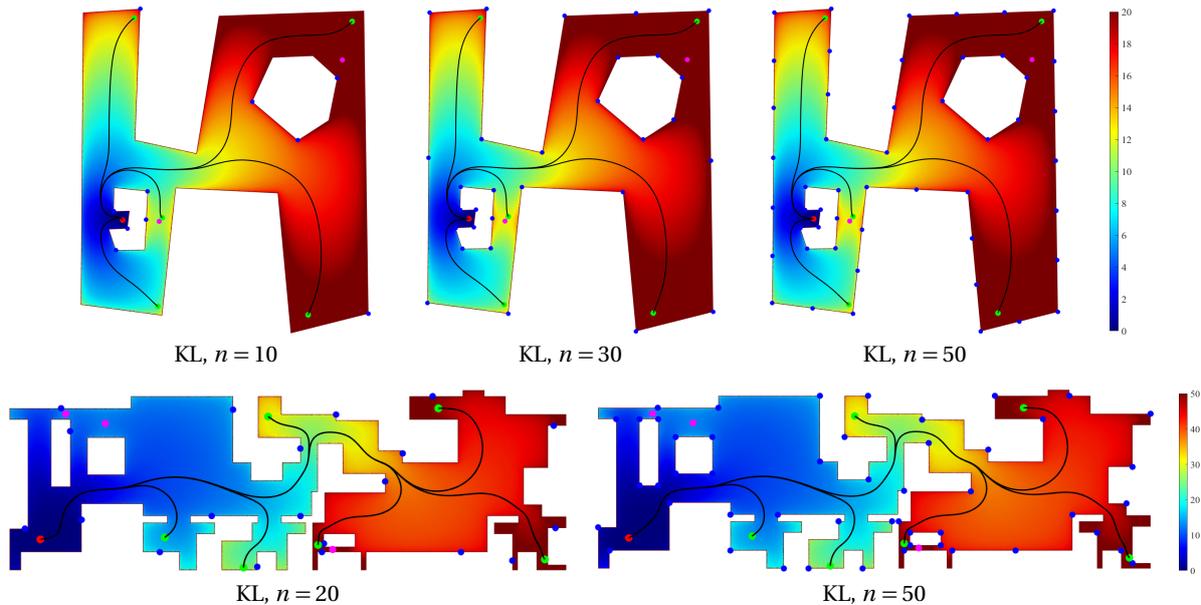
$$\nabla d_f^\Omega(z) = \nabla d_f^D(C(z)) \frac{\partial C}{\partial z}(z).$$

Since the derivative of a conformal map never vanishes, according to Theorem A.1, we have

$$\nabla d_f^\Omega(z) = 0 \iff \nabla d_f^D(C(z)) = 0 \iff C(z) = 0 \iff z = y. \quad \square$$

## 5.1 Examples

Figures 2 and 3 show what happens to the paths in Figure 1 (right) when the reduced  $f$ -divergence distance (4) is used instead of the original (continuous)  $f$ -divergence distance (1), for two choices of  $f$ , the Kullback–Leibler (KL) divergence  $f(x) = -\log x$  and the Hellinger (H) divergence  $f(x) = 2(1 - \sqrt{x})$ . The coordinates are reduced by uniformly partitioning the boundary of the domain. For polygonal boundaries (as in Figure 3), it seems natural to partition according to the polygon edges, namely at least one reduced coordinate per edge. Long edges are further partitioned uniformly until the partition length is less than some threshold. Since reduced coordinates lose the property of being invariant to  $f$ , different paths are obtained for reduced KL and H distances. They are typically not as natural as the original paths, especially when  $n$  is small. Obviously,



**Figure 4:** Same as Figure 2, for multiply-connected polygonal domains. The saddle points of the distance function are marked in magenta.

domain	Chen et al. [4]				$f$ -divergence distance with $n$ reduced coordinates							
	$n$	prepro- cessing	online	memory	$n$	prepro- cessing	online	memory	$n$	prepro- cessing	online	memory
Fig. 2	1,200	14.6	2.38	1,827.9	10	0.903	0.023	15.2	30	0.959	0.061	45.7
Fig. 3	3,805	178.5	6.94	5,814.0	20	0.768	0.057	30.6	30	0.895	0.079	45.7
Fig. 4 top	3,581	117.4	6.84	5,466.8	10	0.810	0.022	15.2	30	0.862	0.074	45.8
Fig. 4 bottom	5,604	303.2	10.6	8,551.7	20	0.710	0.044	30.5	50	0.971	0.113	76.3

**Table 1:** Runtime comparison of Chen et al.’s [4] routing method using a full set of coordinates and our method, using reduced coordinates: times (in seconds) for preprocessing and online computation of the distance from one vertex to all other vertices, and memory cost (in megabytes).

for very large  $n$ , the reduced paths approach the original paths. The same behaviour can be observed for multiply-connected domains, as shown in Figure 4. Note that the distance function has saddle points in this case (typically one per “hole”), but no local minima. In all these examples, we discretize the domains with dense triangulations of 200,000 vertices, and the number of boundary vertices are listed in the second column of Table 1.

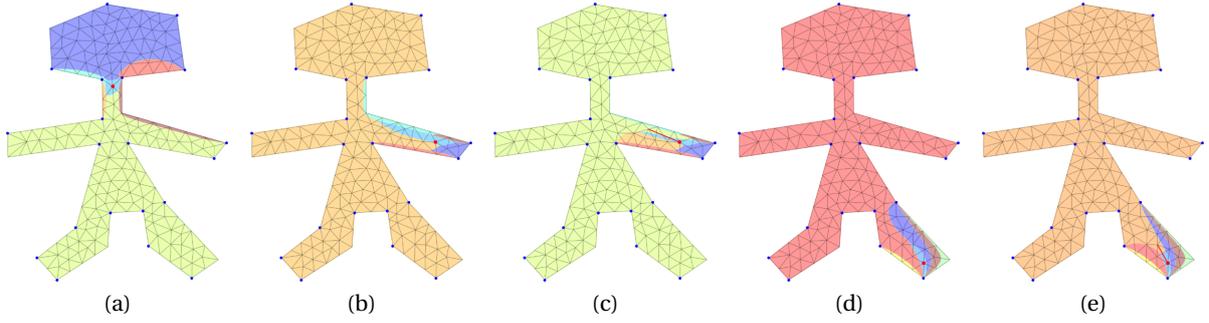
## 6 Computing reduced coordinates

In practice, if  $T$  is a dense triangulation of the domain  $\Omega$  with  $k$  vertices  $x_1, \dots, x_k$ , and  $t_1, \dots, t_n$  is a sequence of points along the boundary of  $T$ , then the reduced coordinates  $\phi_j(x_i)$ ,  $j = 1, \dots, n$  of all vertices can be computed as follows. Let  $A$  be the sparse, symmetric, and positive definite  $k \times k$  matrix with the so-called cotangent weights [18] corresponding to edges of the triangulation, that is, the standard FEM discretization of the Laplace operator for  $T$ . Note that all weights are positive, if a constrained Delaunay triangulation of the domain [9, 21] is used. Further let  $\Phi_j = (\phi_j(x_1), \dots, \phi_j(x_k))$  be the vector of all  $j$ -th reduced coordinates and  $b_j = (b_{j,1}, \dots, b_{j,k})$  be the *binary indicator* of the segment  $[t_j, t_{j+1}]$  of points along the boundary of  $T$  between  $t_j$  and  $t_{j+1}$ ,

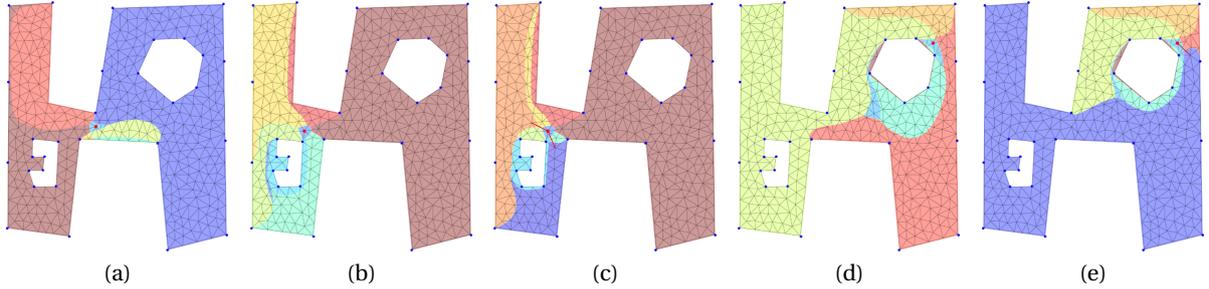
$$b_{j,i} = \begin{cases} 1, & \text{if } x_i \in [t_j, t_{j+1}], \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

We then get all reduced coordinates by solving the  $n$  linear systems

$$A\Phi_j = b_j, \quad j = 1, \dots, n, \quad (6)$$



**Figure 5:** Construction of a greedy routing graph for the reduced KL divergence distance with  $n = 20$  reduced coordinates and  $m = 200$  sites sampled from a simply-connected domain. For the initial Euclidean Delaunay triangulation of the sites (a), it may happen that the local Voronoi cell (light blue) of a site (red) contains other sites (b,d). In this case, we add an edge (red) between these two sites (c,e), until the local Voronoi cells contain no other sites. Note how this affects the local Voronoi cells of the neighbouring sites.



**Figure 6:** Same as Figure 5, for a multiply-connected domain with  $n = 30$  reduced coordinates and  $m = 400$  sites.

which can be done efficiently by pre-factoring  $A$  and performing back-substitution for the different  $b_j$  [19]. Note that the reduced coordinates are computed once for all vertices in a preprocessing stage. These are then used again and again for each “online” routing query (consisting of a pair of source and target points).

## 7 Discrete routing graph

We now show how to make path generation with reduced  $f$ -divergence distances even more practical. To this end, let  $S$  be a finite and sparse set of  $m$  sites in the domain  $\Omega$ , and  $G$  be a graph on  $S$ , and consider paths running along the edges of  $G$ . To be able to mimic the gradient-descent idea in this discrete setting, we need a graph  $G$  that supports *greedy routing* to any  $y \in S$ ,

$$\forall s \in S \setminus \{y\} : \exists r \in N_G(s) : d_f(y, r) < d_f(y, s), \quad (7)$$

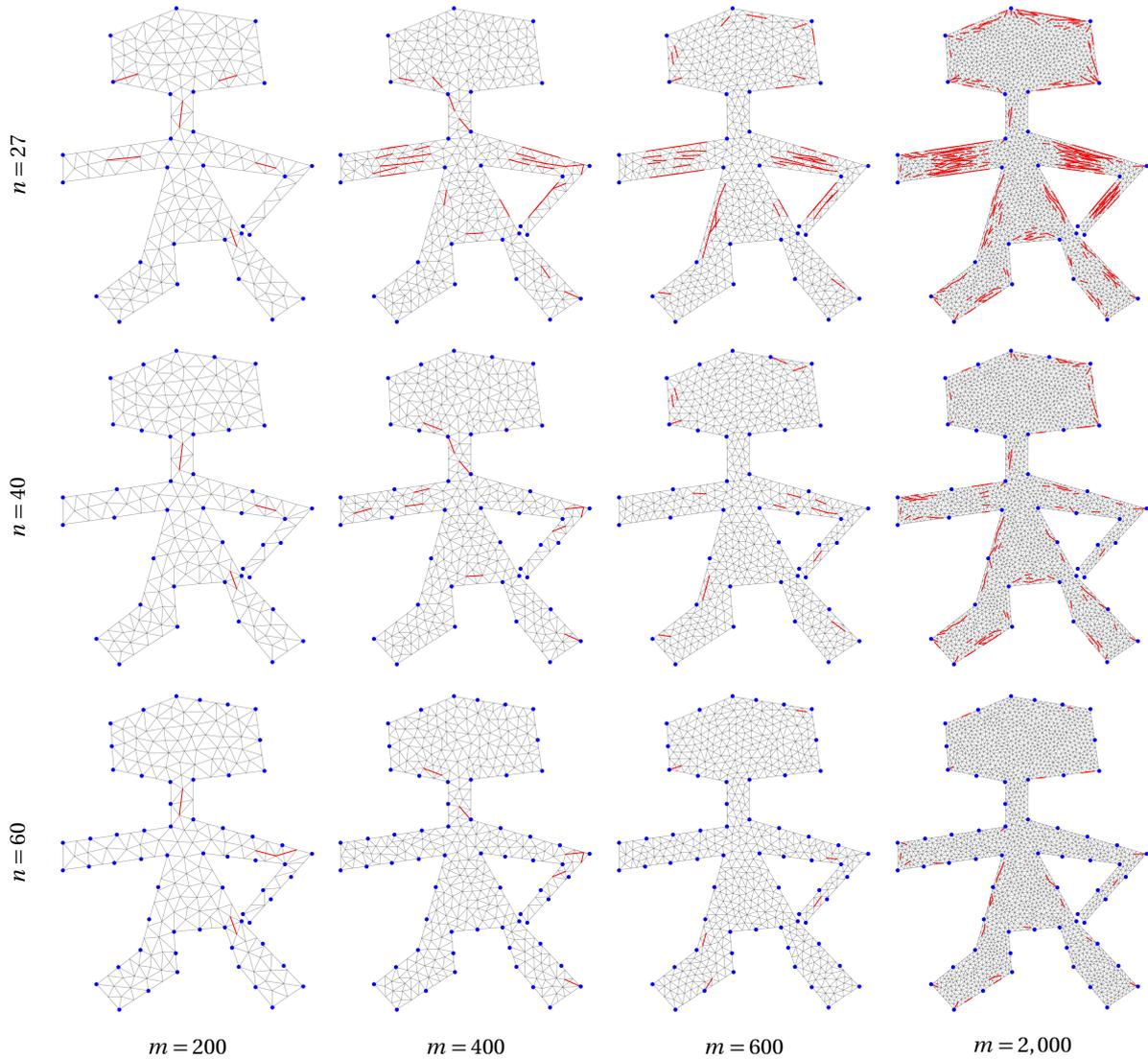
where  $N_G(s)$  is the set of neighbours of  $s$  in  $G$ .

To achieve this, and inspired by [2], we follow the logic of Bose and Morin [3], who show that the *Delaunay triangulation* [9] of  $S$  supports greedy routing on the convex hull of  $S$ , using the Euclidean distance between points in the plane. The reason for the Delaunay triangulation to have this property is because it is the dual of the Euclidean *Voronoi diagram* of  $S$ . That is, two sites are connected by an edge if and only if their two corresponding Voronoi cells share a common edge. The proof that greedy routing works relies on the fact that the Euclidean distance is a metric and its Voronoi cells are convex polygons. Since the Voronoi diagram of  $S$  using a reduced  $f$ -divergence distance is more complicated, the condition must be modified, requiring the concept of a *local Voronoi cell* of a site in a graph.

**Definition 3.** The local Voronoi cell of  $s \in S$  relative to the graph  $G$  is the set of all points in  $\Omega$  for which  $s$  is closer than any of the neighbours of  $s$  in  $G$ ,

$$LV_G(s) = \{z \in \Omega : d_f(z, s) < d_f(z, r) \forall r \in N_G(s)\}.$$

The *greedy routing property* now guarantees that (7) is satisfied.



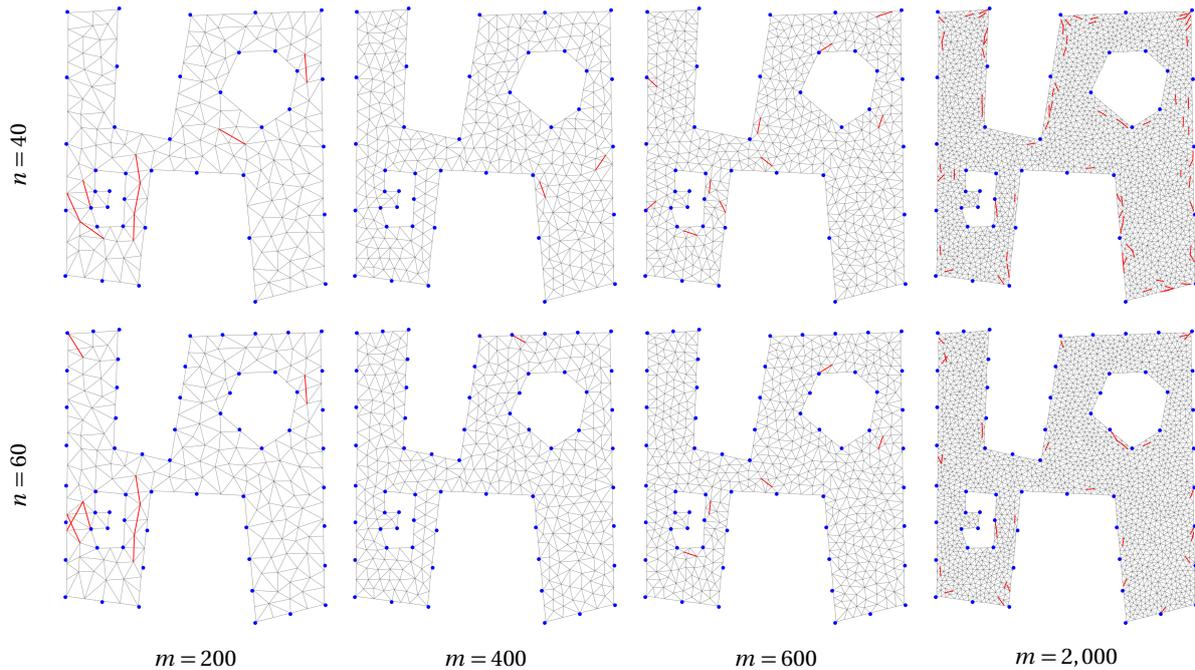
**Figure 7:** Discrete greedy routing graph using the KL divergence distance with  $n$  reduced coordinates and  $m$  sites on a simply connected domain for different values of  $n$  and  $m$ . The greedy routing property is obtained after augmenting the initial constrained Delaunay triangulation (black) with additional edges (red). Note how the number of augmented edges decreases with increasing  $n$ .

**Definition 4.** The graph  $G$  has the greedy routing property, if for every site  $s \in S$ ,  $LV_G(s)$  does not contain any site other than  $s$ ,

$$\forall s, t \in S: \quad t \in LV_G(s) \iff t = s.$$

Constructing a graph  $G$  on  $S$  having the greedy routing property is not as straightforward as it seems. It is not sufficient to merely take the dual of the Voronoi diagram of  $S$ , because the local Voronoi cells of the reduced  $f$ -divergence distance may have irregular structure. They need not even be connected and can have so-called *orphan cells*, and the reduced  $f$ -divergence distance may be asymmetric. This is especially true when the number  $n$  of coordinates is small. On the other hand, the clique graph, which connects all sites with each other, obviously has the greedy routing property, but this is a gross overkill, as we would like to have a graph that is as sparse as possible, with edges as short as possible. A planar graph would be the most desirable.

In order to build a greedy routing graph  $G$ , given  $S$  and  $d_f$ , we propose an incremental algorithm. Starting with the Euclidean Delaunay triangulation, constrained to a polygonal approximation of the domain, we *augment* this initial graph with additional edges until it becomes greedy. Given a site  $s \in S$ , edges are added to  $G$  between  $s$  and other sites, until  $LV_G(s)$  contains only  $s$ . By definition, each addition of an edge shrinks



**Figure 8:** Same as Figure 7, for a multiply-connected domain.

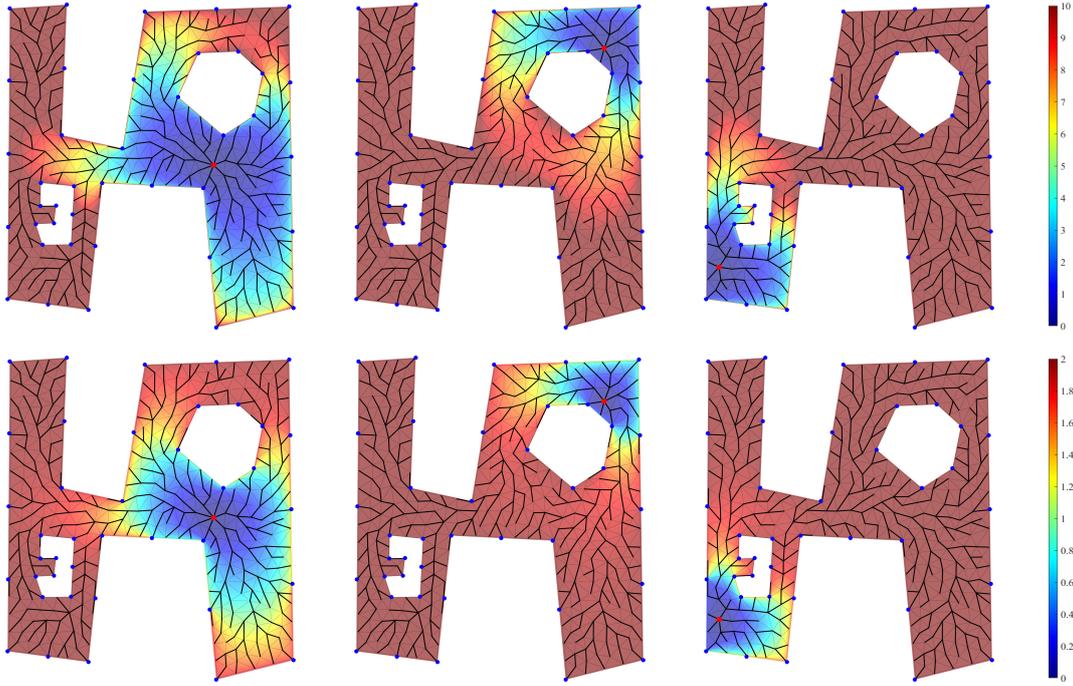
domain	$n$	$m = 200$		$m = 400$		$m = 600$		$m = 2,000$	
		distances	augmenting	distances	augmenting	distances	augmenting	distances	augmenting
Fig. 7	27	0.033	0.0149	0.090	0.0084	0.191	0.0138	1.344	0.1247
	40	0.059	0.0142	0.111	0.0082	0.243	0.0109	1.732	0.0837
	60	0.078	0.0118	0.147	0.0089	0.319	0.0092	2.472	0.0757
Fig. 8	40	0.050	0.0152	0.103	0.0063	0.234	0.0117	1.707	0.0808
	60	0.081	0.0165	0.177	0.0056	0.329	0.0122	2.675	0.0783

**Table 2:** Runtime for generating discrete routing graphs. For each configuration of  $m$  and  $n$ , we list the times (in seconds) for computing the pairwise reduced  $f$ -divergence distances between the  $m$  sites and for augmenting the constrained Delaunay triangulation.

$LV_G(s)$ . Obviously, this procedure eventually terminates when the worst case of  $s$  being connected to all other sites is obtained. A good heuristic is to add edges between  $s$  and other sites in order of increasing Euclidean distance to  $s$ . We call the resulting greedy graph the *augmented Delaunay triangulation*. In practice, no Voronoi diagrams are computed, and the only data structure required to support the graph construction algorithm is the  $m \times m$  matrix of pairwise reduced  $f$ -divergence distances between the  $m$  sites, sampled from the dense underlying triangulation that we use to precompute the coordinates. Note that this matrix is not symmetric if  $d_f$  is not symmetric.

## 8 Experimental results

We implemented our method in *Matlab R2018a* on a PC with Intel Xeon (6 cores) CPU E5-2643 v4 3.40Ghz and 32GB RAM. For all the examples throughout the paper, the underlying constrained Delaunay triangulation  $T$  of the domain on which the  $n$  coordinates are computed contains  $k = 200,000$  vertices. Table 1 lists the runtimes for computing  $n$  reduced coordinates and  $f$ -divergence distances (from one vertex to all other vertices of  $T$ ), and the memory footprint for storing the reduced coordinates, in comparison to Chen et al. [4], where  $n$  is the number of boundary vertices of  $T$ . We observe that our method is faster both in preprocessing and online distance computation, and the memory requirement is much lower. Note that the system matrices involved in the two approaches are identical, yet the preprocessing takes significant more time for Chen et al. [4], since they need to solve for significantly more right hand sides in order to get a good approximation of the harmonic measure.

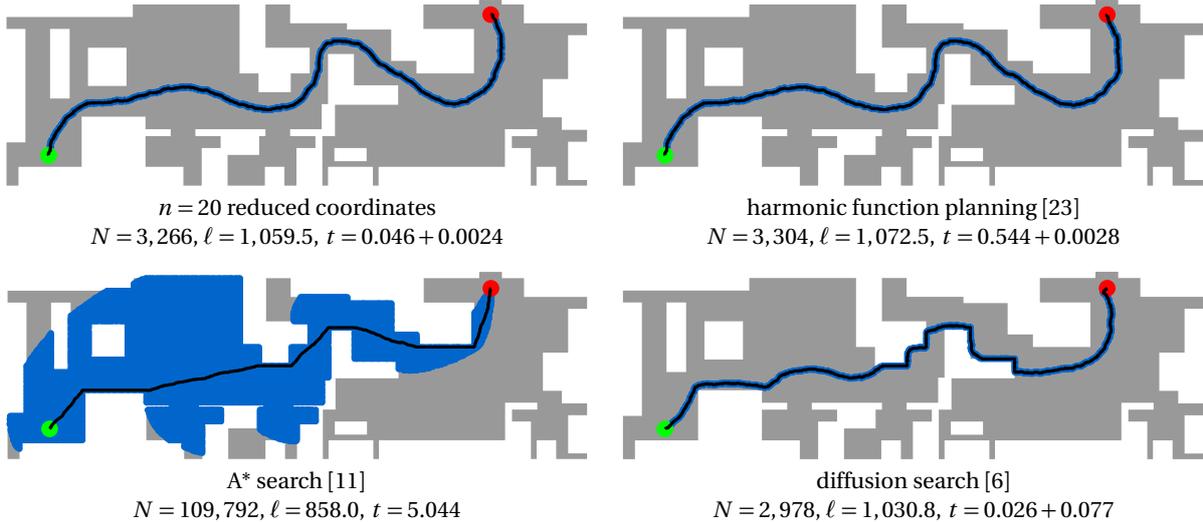


**Figure 9:** Routing path tree (thick black edges) for three different target sites (red) for the multiply-connected domain and graph of Figure 8 for  $n = 40$  reduced coordinates and  $m = 400$  sites using the KL (top) and the H (bottom) divergence distance. The domain is colour-coded according to the  $f$ -divergence distance of the reduced coordinates to the target site. Note the subtle differences between the two rows.

Figures 5 and 6 show parts of the greedy routing graphs generated by our algorithm on a few sites sampled in two different domains with 3,805 and 3,581 boundary vertices, respectively. The black edges are the initial constrained Delaunay triangulation, and the red edges are those augmented by our algorithm for three select sites to obtain the greedy routing property.

Figures 7 and 8 show the greedy augmented Delaunay triangulation on three increasing sets of sites with increasing sets of reduced coordinates using a similar colouring of the edges. As before, the boundary polygon was naturally partitioned to its edges, and then each edge further partitioned to segments until the length of each segment is below a predefined threshold. The results demonstrate how the Delaunay triangulation is already very close to greedy, requiring only a slight augmentation, for large values of  $n$ . Table 2 lists the runtimes for generating the greedy routing graphs of the augmented Delaunay triangulations. Note that all this is done in a preprocessing stage. Figure 9 shows routing trees for three target vertices on the resulting graph for two different distance functions on a multiply-connected shape.

The idea of using distance functions for path planning in robotic applications promises to remain a popular methodology. Indeed, a recent paper of Chen et al. [6] uses diffusion distances for this objective. Similarly to our divergence distances, diffusion distances may be computed efficiently using a small set of *embedding coordinates* derived from the eigenvectors of the graph Laplacian having the largest eigenvalues, computed in a preprocessing stage. However, this distance function is not void of local minima, so the *best-first-search* algorithm is used to generate paths on a discrete network, incurring overhead by examining many neighbouring nodes in the vicinity of the path and maintaining a priority queue. Note that if the distance function is void of local minima, best-first-search reduces to gradient descent and a priority queue is not required. In principle, we could have adopted a similar approach and applied best-first-search instead of gradient descent, avoiding the need to augment the graph. However, as our experimental results indicate, the augmentation required is minimal for sparse triangulation and most of the time unnecessary for sufficiently densely triangulated domains, due to the fact that the continuous version of our reduced divergence distance is greedy, as compared to the continuous diffusion distance, which does not have this desirable property. Thus we believe it is preferable to pay the small price of augmenting the graph when necessary in the preprocessing stage and benefiting significantly from a very simple routing algorithm in the online query stage. Figure 10 compares our method with harmonic potential, A\* search [11] (with the Euclidean distance as a heuristic function), and diffusion search [6] (based on 18 eigenvectors) for the laboratory map from



**Figure 10:** Comparison of different path planning algorithms. The black polyline marks the path between the source (green) and target (red) found by each algorithm. To determine this path, each algorithm needs to compute the distance function for the  $N$  vertices in the blue region around the path. We report  $N$ , the length  $\ell$  of each path, and the planning time  $t$  (in seconds). Note that  $t$  does not include the preprocessing time (0.737 s for our method and 93 s for diffusion search), but for harmonic function planning we include the linear solve time of 0.544 s, as it has to be computed online after the routing target is selected, and for our method and diffusion search we include the 0.046 s and 0.026 s that it takes, respectively, to compute the distance from the target to every other vertex.

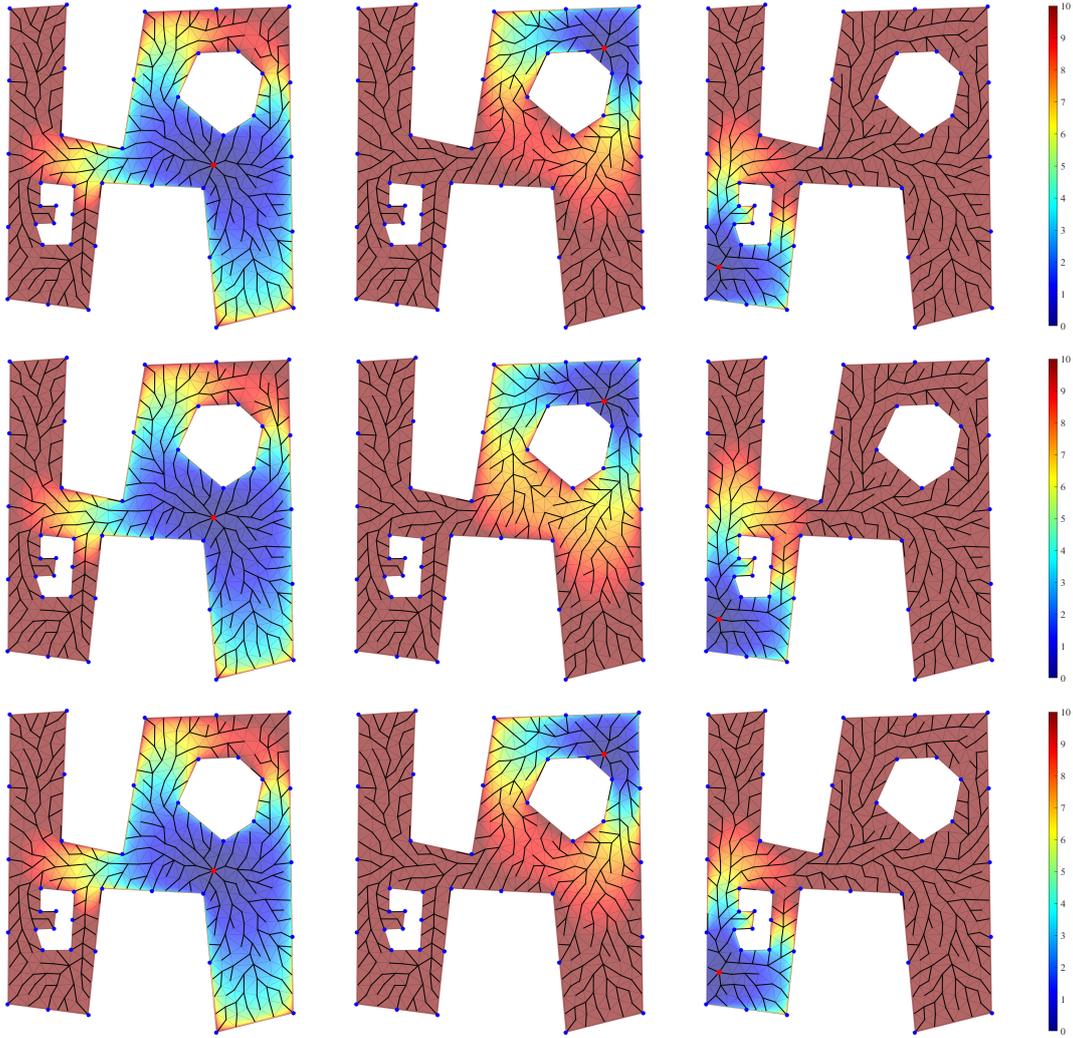
Wray et al. [23]. For harmonic function planning, we directly compute the potential function by a sparse linear solver instead of its logarithm as in [23], as it is more efficient and provides sufficient accuracy for greedy routing in this case. For diffusion search, we used a time step of 200 times the width of the domain, because the factor of 50 suggested in [6] gives significantly worse values of  $N$  and  $t$ .

## 9 Conclusion and discussion

We have described a practical method for generating paths between pairs of points in planar domains by following the negative gradient of a suitable distance function. The key is to use a discrete version of the  $f$ -divergence distance, based on reduced coordinate vectors. Our reduced coordinate vectors are the harmonic measures of a partition of the boundary into  $n$  segments, which are just the inner products of the Poisson kernel with a “box” indicator function for each segment. We speculate that the Divergence Gradient Theorem 1 also holds for other sets of coordinates, derived from an inner product of the Poisson kernel with more sophisticated basis functions, such as the piecewise linear “tent” function over two adjacent boundary segments. This is the method used to construct *harmonic barycentric coordinates* on a polygonal domain [12]. Another possible choice are Gaussian basis functions over the boundary, as long as they are not too narrow or too wide. These variants of reduced coordinates may be computed similarly to the reduced coordinates used in this paper merely by changing the right hand sides  $b_j$  of the linear equations in (6) to something more sophisticated than the binary vector in (5). Figure 11 shows the gradient-descent trees generated by these coordinates using the KL divergence distance function.

The proof of our Divergence Gradient Theorem 1 holds only for bounded and simply-connected domains, as it relies on conformal reduction to the canonical case of a unit disk with the target at the origin. We do not have a proof for the absence of local minima for multiply-connected domains, where the holes in the domain correspond to obstacles in a real-world scenario, but speculate that the theorem still holds. This is supported by all our experiments, as well as a recent result on the  $f$ -divergence distance for the  $\chi^2$ -divergence  $f(x) = x^2 - 1$  in the continuous setting [5]. Note that for multiply-connected domains, the distance function may contain saddle points, where the continuous gradient vanishes, but as these are not local minima, they are not fatal for path generation.

Our experimental results also indicate that as the number of reduced coordinates increases, the augmentation of the Delaunay triangulation of the sites decreases. We wonder if there exists a condition, possibly



**Figure 11:** Same as in Figure 9, for  $n = 40$  reduced coordinates and  $m = 400$  sites using the KL divergence distance and reduced coordinates based on box (top), tent (middle), and Gaussian (bottom) basis functions over the boundary.

dependent on  $n$ , the number of coordinates, and  $m$ , the number of sites, which guarantees that the Delaunay triangulation is greedy in its own right. Alternatively, under which conditions does there exist a planar graph, possibly a non-Delaunay triangulation, of the sites, which is greedy? This may well be the dual of the Voronoi diagram for  $d_f$ .

It is worth noting that the harmonic potential with pole at the target point  $y \in \Omega$ , which has traditionally been used for path planning, can also be expressed as a boundary integral once the Poisson kernel of the domain has been computed:

$$D(y, z) = -\log|z - y| + \oint_{\partial\Omega} P(t, z) \log|t - y| dt.$$

So theoretically, we could compute this potential quite efficiently at any given point  $z \in \Omega$  as a boundary integral (once the Poisson kernels are computed in a preprocessing stage), instead of solving a large system of linear equations. Thus the natural question arises: Why is it better to compute the divergence distance as a boundary integral and not the harmonic potential?

One reason is that, in the discrete case, the harmonic potential computed as a boundary integral is not discrete harmonic (namely, does not satisfy the discrete Laplace equation), thus may contain local minima. This is in contrast to explicitly solving the discrete Laplace equation. So, to use it in practice, we will have to apply graph augmentation, which may be more excessive than in the case of the augmentation of the graph by divergence distance.

More importantly, the harmonic potential is much less attractive because we cannot apply coordinate reduction to reduce the complexity of the boundary integral, which we can do for the divergence distance, thanks to Theorem 1.

## References

- [1] S. Axler, P. Bourdon, and W. Ramey. *Harmonic Function Theory*, volume 137 of *Graduate Texts in Mathematics*. Springer, New York, 2nd edition, 2001.
- [2] M. Ben-Chen, S. J. Gortler, C. Gotsman, and C. Wormser. Distributed computation of virtual coordinates for greedy routing in sensor networks. *Discrete Applied Mathematics*, 159(7):544–560, Apr. 2011.
- [3] P. Bose and P. Morin. Online routing in triangulations. *SIAM Journal on Computing*, 33(4):937–951, 2004.
- [4] R. Chen, C. Gotsman, and K. Hormann. Path-planning with divergence-based distance functions. arXiv:1708.05855 [cs.RO], 2017. <https://arxiv.org/pdf/1708.05855>.
- [5] R. Chen, C. Gotsman, and K. Hormann. On divergence-based distance functions for multiply-connected domains. arXiv:1801.07099 [math.CV], 2018. <https://arxiv.org/pdf/1801.07099>.
- [6] Y. F. Chen, S.-Y. Liu, M. Liu, J. Miller, and J. P. How. Motion planning with diffusion maps. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1423–1430, Daejeon, Oct. 2016.
- [7] C. I. Connolly and R. A. Grupen. The applications of harmonic functions to robotics. *Journal of Field Robotics*, 10(7):931–946, Oct. 1993.
- [8] I. Csizsár. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2:299–318, 1967.
- [9] M. de Berg, O. Cheong, M. van Kreveld, and M. Overmars. *Computational Geometry: Algorithms and Applications*. Springer, Berlin, 3rd edition, 2008.
- [10] J. B. Garnett and D. E. Marshall. *Harmonic Measure*, volume 2 of *New Mathematical Monographs*. Cambridge University Press, New York, 2005.
- [11] P. E. Hart, N. J. Nilsson, and B. Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968.
- [12] P. Joshi, M. Meyer, T. DeRose, B. Green, and T. Sanocki. Harmonic coordinates for character articulation. *ACM Transactions on Graphics*, 26(3):Article 71, 9 pages, July 2007.
- [13] O. Khatib. Real-time obstacle avoidance for manipulators and mobile robots. *International Journal of Robotics Research*, 5(1):90–98, 1986.
- [14] J.-O. Kim and P. K. Khosla. Real-time obstacle avoidance using harmonic potential functions. *IEEE Transactions on Robotics and Automation*, 8(3):338–349, June 1992.
- [15] Y. Koren and J. Borenstein. Potential field methods and their inherent limitations for mobile robot navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1398–1404, Sacramento, Apr. 1991.
- [16] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, Mar. 1951.
- [17] F. Liese and I. Vajda. On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory*, 52(10):4394–4412, Oct. 2006.
- [18] U. Pinkall and K. Polthier. Computing discrete minimal surfaces and their conjugates. *Experimental Mathematics*, 2(1):15–36, 1993.
- [19] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, New York, 3rd edition, 2007.
- [20] E. Rimon and D. E. Koditschek. Exact robot navigation using artificial potential functions. *IEEE Transactions on Robotics and Automation*, 8(5):501–518, Oct. 1992.
- [21] J. R. Shewchuk. Triangle: Engineering a 2D quality mesh generator and Delaunay triangulator. In M. C. Lin and D. Manocha, editors, *Applied Computational Geometry. Towards Geometric Engineering*, volume 1148 of *Lecture Notes in Computer Science*, pages 203–222. Springer, Berlin, 1996.
- [22] O. Souissi, R. Benatitallah, D. Duvivier, A. Artiba, N. Belanger, and P. Feyzeau. Path planning: A 2013 survey. In *Proceedings of the IEEE International Conference on Industrial Engineering and Systems Management*, pages 849–856, Rabat, Oct. 2013.
- [23] K. H. Wray, D. Ruiken, R. A. Grupen, and S. Zilberstein. Log-space harmonic function path planning. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1511–1516, Daejeon, Oct. 2016.

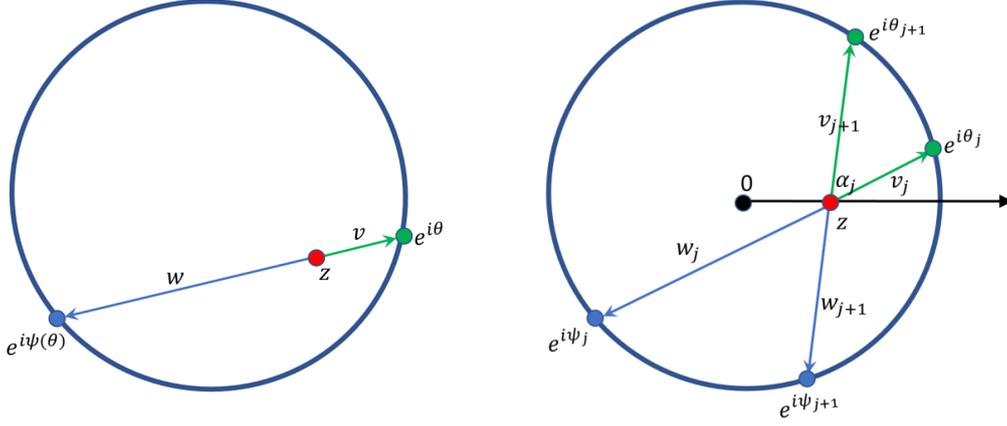


Figure 12: Notation used in Sections A.1 (left) and A.2 (right).

## A Appendix

In this appendix, we prove Theorem 1 for the special case of the unit disk domain with the target point at the origin.

### A.1 Some circle geometry

In what follows we use complex number algebra in the plane. Let  $z$  be some point in the unit disk  $D$ . As shown in Figure 12 (left), we denote by  $\psi(\theta) \in (-\pi, \pi]$  the *antipode* of  $\theta \in (-\pi, \pi]$  relative to  $z$ , that is,  $e^{i\psi(\theta)}$  is the intersection of the chord through  $e^{i\theta}$  and  $z$  with the unit circle. We further let  $v = e^{i\theta} - z$  and  $w = e^{i\psi(\theta)} - z$ .

**Lemma A.1.** For any  $\theta \in (-\pi, \pi]$ ,

$$\frac{1}{\bar{v}} = -\frac{w}{1-|z|^2}.$$

*Proof.* By the intersecting chords theorem,  $|v||w| = 1 - |z|^2$ . Since  $v$  and  $w$  are collinear, this implies  $-\bar{v}w = |v||w| = 1 - |z|^2$ .  $\square$

**Lemma A.2.** For any  $\theta \in (-\pi, \pi]$ ,

$$\psi(\theta) = -i \log\left(-\frac{v}{\bar{v}} e^{-i\theta}\right).$$

*Proof.* By the definition of  $\psi(\theta)$ , we have  $\psi(\theta) = -i \log(w + z)$ , and it follows from Lemma A.1 that  $w + z = -v e^{-i\theta} / \bar{v}$ .  $\square$

Note that for  $z = 0$ , we have  $\psi(\theta) = \pi + \theta$ , as expected.

### A.2 Reduced coordinates

Given a partition  $-\pi < \theta_1 < \dots < \theta_n \leq \pi$  of the unit circle, the  $j$ -th reduced coordinate of  $z \in D$  is the harmonic measure of the arc  $(\theta_j, \theta_{j+1})$  at  $z$ , that is,

$$\phi_j(z) = \int_{\theta_j}^{\theta_{j+1}} P^D(\theta, z) d\theta = \frac{1}{2\pi} (2\alpha_j(z) - (\theta_{j+1} - \theta_j)), \quad (8)$$

where  $P^D(\theta, z)$  is the Poisson kernel of the unit disk,

$$P^D(\theta, z) = \frac{1}{2\pi} \frac{1 - |z|^2}{|z - e^{i\theta}|^2},$$

and  $\alpha_j(z)$  denotes the angle the arc forms with  $z$ , as shown in Figure 12 (right). Note that since  $e^{i\theta}$  is a  $2\pi$ -periodic function, we may use  $\theta_{n+1} = \theta_1 + 2\pi$ , so that  $\theta_{j+1} - \theta_j$  always gives the positive length of the arc, and then  $\phi_j(z) \geq 0$  and  $\sum_j \phi_j(z) = 1$ .

**Lemma A.3.** The harmonic measure in (8) can also be expressed as

$$\phi_j(z) = \frac{1}{2\pi}(\psi_{j+1} - \psi_j), \quad (9)$$

where  $\psi_j = \psi(\theta_j)$  is the antipode of  $\theta_j$  relative to  $z$ .

*Proof.* By the intersecting chords theorem,  $\theta_{j+1} - \theta_j + \psi_{j+1} - \psi_j = 2\alpha_j(z)$ , and the statement then follows directly from (8).  $\square$

**Lemma A.4.** The gradient (by  $z$ ) of  $\phi_j(z)$  is

$$\nabla \phi_j(z) = \frac{i}{\pi(1-|z|^2)}(e^{i\psi_{j+1}} - e^{i\psi_j}).$$

*Proof.* By Lemma A.3 and viewing  $\psi_j$  as a function of  $z$ , we get

$$\nabla \phi_j(z) = \frac{1}{2\pi} \nabla(\psi_{j+1}(z) - \psi_j(z)).$$

Using Lemma A.2 and the complex form of the gradient then gives

$$\nabla \psi_j(z) = -i \nabla \log\left(\frac{-v_j}{\bar{v}_j}\right) = -2i \left(\frac{\partial}{\partial z} \log\left(\frac{-v_j}{\bar{v}_j}\right)\right) = -\frac{2i}{\bar{v}_j},$$

so

$$\nabla \phi_j(z) = -\frac{i}{\pi} \left(\frac{1}{\bar{v}_{j+1}} - \frac{1}{\bar{v}_j}\right).$$

Applying Lemma A.1, we finally get

$$\nabla \phi_j(z) = \frac{i}{\pi(1-|z|^2)}(w_{j+1} - w_j) = \frac{i}{\pi(1-|z|^2)}(e^{i\psi_{j+1}} - e^{i\psi_j}).$$

$\square$

### A.3 Reduced $f$ -divergence distance

Given a partition  $-\pi < \theta_1 < \dots < \theta_n \leq \pi$  of the unit circle and a strictly convex function  $f$ , the reduced  $f$ -divergence distance in (4) from  $z \in D$  to  $y = 0$  can be written as

$$d_f(z) = d_f(0, z) = \sum_{j=1}^n \phi_j(0) f\left(\frac{\phi_j(z)}{\phi_j(0)}\right), \quad (10)$$

with  $\phi_j(z)$  in (9). By the chain rule and Lemma A.4,

$$\nabla d_f(z) = \sum_{j=1}^n \phi_j(0) \nabla f\left(\frac{\phi_j(z)}{\phi_j(0)}\right) = \sum_{j=1}^n f'\left(\frac{\phi_j(z)}{\phi_j(0)}\right) \nabla \phi_j(z) = \frac{i}{\pi(1-|z|^2)} \sum_{j=1}^n f'\left(\frac{\psi_j - \psi_{j+1}}{\theta_j - \theta_{j+1}}\right) (e^{i\psi_{j+1}} - e^{i\psi_j}).$$

Without loss of generality, let us now assume that  $z$  is on the positive  $x$ -axis, so that  $z = \bar{z} = |z|$ . Any other case can be reduced to this setting by a simple rotation of the plane.

**Lemma A.5.** The function  $g(\psi) = f'(1/\theta'(\psi))$  is strictly increasing for  $\psi \in (0, \pi)$  and strictly decreasing for  $\psi \in (-\pi, 0)$ , if  $z \neq 0$ , and constant, if  $z = 0$ .

*Proof.* It follows from Lemma A.2 that

$$\psi(\theta) = -i \log\left(-\frac{e^{i\theta} - z}{e^{-i\theta} - z} e^{-i\theta}\right),$$

hence

$$\psi'(\theta) = \frac{e^{i\theta}}{e^{i\theta} - z} + \frac{e^{-i\theta}}{e^{-i\theta} - z} - 1 = \frac{z}{e^{i\theta} - z} + \frac{z}{e^{-i\theta} - z} + 1 = 2z \operatorname{Re}\left(\frac{1}{e^{i\theta} - z}\right) + 1$$

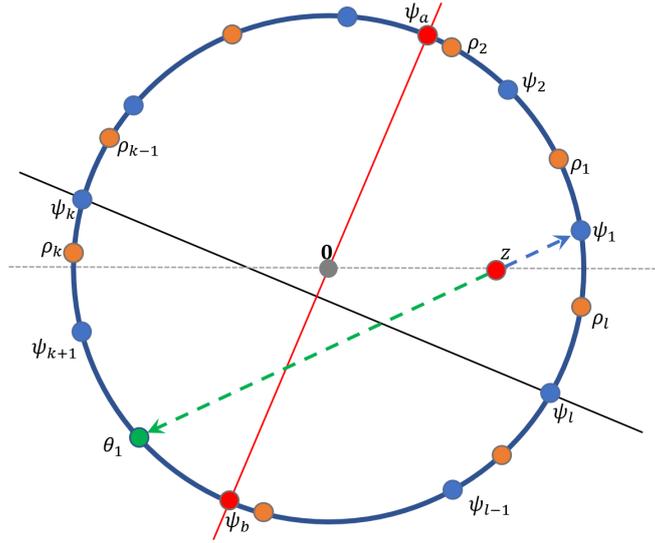


Figure 13: Notation used in the proof of Theorem A.1.

and

$$\psi''(\theta) = -\frac{iz e^{i\theta}}{(e^{i\theta} - z)^2} + \frac{iz e^{-i\theta}}{(e^{-i\theta} - z)^2} = 2z \operatorname{Im} \left( \frac{e^{i\theta}}{(e^{i\theta} - z)^2} \right) = \frac{-2z(1-z^2)}{|e^{i\theta} - z|^4} \sin \theta \begin{cases} = 0, & \text{if } z = 0, \\ < 0, & \text{if } z \neq 0 \text{ and } \theta \in (0, \pi), \\ > 0, & \text{if } z \neq 0 \text{ and } \theta \in (-\pi, 0). \end{cases}$$

Since  $\theta(\psi)$  and  $\psi(\theta)$  are antipodes of each other, they have the same behaviour, and therefore

$$\theta''(\psi) \begin{cases} = 0, & \text{if } z = 0, \\ < 0, & \text{if } z \neq 0 \text{ and } \psi \in (0, \pi), \\ > 0, & \text{if } z \neq 0 \text{ and } \psi \in (-\pi, 0). \end{cases} \quad (11)$$

Applying the chain rule to  $g(\psi)$ , we find that

$$g'(\psi) = -f'' \left( \frac{1}{\theta'(\psi)} \right) \frac{\theta''(\psi)}{\theta'(\psi)^2},$$

and as  $f'' > 0$ , since  $f$  is strictly convex, we conclude that the sign of  $g'$  is opposite to the sign of  $\theta''(\psi)$ . The statement then follows from (11).  $\square$

Note that both  $g(\psi)$  and  $\psi'(\theta)$  are even functions.

**Theorem A.1** (Divergence Gradient Theorem for the unit disk). For any partition  $-\pi < \theta_1 < \dots < \theta_n \leq \pi$  of the unit circle with  $n \geq 3$ , the reduced  $f$ -divergence distance to the origin in (10) satisfies

$$\nabla d_f(z) = 0 \quad \iff \quad z = 0.$$

*Proof.* For each arc  $(\theta_j, \theta_{j+1})$ , the mean-value theorem states that there exists some  $\tau_j \in (\theta_j, \theta_{j+1})$ , such that  $\psi'(\tau_j) = \frac{\psi_j - \psi_{j+1}}{\theta_j - \theta_{j+1}}$ . Now let  $\rho_j = \psi(\tau_j) \in (\psi_j, \psi_{j+1}) \subset (-\pi, \pi]$  and define the piecewise constant periodic function  $w(\psi) = g(\rho_j)$ ,  $\psi \in (\psi_j, \psi_{j+1}]$ . Obviously there exist  $k$  and  $l$  such that  $|\rho_k| \leq |\rho_j| \leq |\rho_l|$  for all  $j$ , that is, the indices  $k$  and  $l$  correspond to the leftmost and rightmost  $\rho$ 's, respectively. In particular, as shown in Figure 13,  $l$  is either 1,  $n-1$ , or  $n$ . Since  $g$  is strictly increasing in  $(0, \pi)$  and strictly decreasing in  $(-\pi, 0)$  by Lemma A.5,  $w(\psi)$  is monotonically (but not strictly) increasing in  $(\psi_l, \psi_k)$  (if  $\psi_k > 0$ , otherwise we use  $(\psi_l, \psi_k + 2\pi]$  and the same principle applies below), and monotonically decreasing in  $(\psi_k, \psi_l)$  by construction. Now let  $[e^{i\psi_a}, e^{i\psi_b}]$  be the diameter orthogonal to  $[e^{i\psi_k}, e^{i\psi_l}]$ . The gradient of  $d_f$  can then be written as

$$\nabla d_f(z) = \frac{i}{\pi(1-|z|^2)} \oint_C w(\psi) de^{i\psi} = \frac{i}{\pi(1-|z|^2)} \left( \int_{\psi_l}^{\psi_k} w(\psi) de^{i\psi} + \int_{\psi_k}^{\psi_l} w(\psi) de^{i\psi} \right) = \nabla d_f^1 + \nabla d_f^2$$

and the projection of  $\nabla d_f^1$  onto  $[e^{i\psi_l}, e^{i\psi_k}]$  be expressed as

$$\begin{aligned}
\langle i e^{i\psi_a}, \nabla d_f^1 \rangle &= \left\langle i e^{i\psi_a}, \frac{i}{\pi(1-|z|^2)} \int_{\psi_l}^{\psi_k} w(\psi) d e^{i\psi} \right\rangle \\
&= \frac{-1}{\pi(1-|z|^2)} \left\langle e^{i\psi_a}, \int_{\psi_l}^{\psi_k} i w(\psi) e^{i\psi} d\psi \right\rangle \\
&= \frac{-1}{\pi(1-|z|^2)} \int_{\psi_l}^{\psi_k} \operatorname{Re}(-i w(\psi) e^{i(\psi_a - \psi)}) d\psi \\
&= \frac{-1}{\pi(1-|z|^2)} \int_{\psi_l}^{\psi_a} w(\psi) \sin(\psi_a - \psi) d\psi + \frac{-1}{\pi(1-|z|^2)} \int_{\psi_a}^{\psi_k} w(\psi) \sin(\psi_a - \psi) d\psi \\
&= \frac{-1}{\pi(1-|z|^2)} \int_{\psi_l}^{\psi_a} (w(\psi) - w(2\psi_a - \psi)) \sin(\psi_a - \psi) d\psi.
\end{aligned}$$

Now observe that for  $\psi \in (\psi_l, \psi_a)$  we have

$$2\psi_a - \psi \in (\psi_a, 2\psi_a - \psi_l) = (\psi_a, \psi_k),$$

and therefore, since  $w$  is monotonically increasing in  $(\psi_l, \psi_k]$ ,

$$w(\psi) \leq w(2\psi_a - \psi).$$

Moreover,  $\sin(\psi_a - \psi) > 0$ , because  $\psi_a - \psi \in (0, \psi_a - \psi_l) \subset (0, \pi)$ . Overall, we conclude that

$$\int_{\psi_l}^{\psi_a} (w(\psi) - w(2\psi_a - \psi)) \sin(\psi_a - \psi) d\psi \leq 0,$$

hence  $\langle i e^{i\psi_a}, \nabla d_f^1 \rangle \geq 0$ , with equality if and only if  $w(\psi)$  is constant on  $(\psi_l, \psi_k)$ , which happens only in the case  $k = l + 1$ , when the harmonic measure of the arc  $(\theta_l, \theta_k)$  reduces to a single coordinate. Similarly, the projection of  $\nabla d_f^2$  onto  $[e^{i\psi_l}, e^{i\psi_k}]$  satisfies

$$\langle i e^{i\psi_a}, \nabla d_f^2 \rangle = \left\langle i e^{i\psi_a}, \frac{i}{\pi(1-|z|^2)} \int_{\psi_k}^{\psi_l} w(\psi) d e^{i\psi} \right\rangle \geq 0,$$

again with equality if and only if  $w(\psi)$  is constant on  $(\psi_k, \psi_l)$ , or equivalently  $l = k + 1$ . Overall, we conclude  $\langle i e^{i\psi_a}, \nabla d_f(z) \rangle > 0$ , hence  $\nabla d_f(z) \neq 0$ , as long as  $n > 2$ . If  $n = 2$ , then  $w$  is constant over the two integral intervals, hence  $\langle i e^{i\psi_a}, \nabla d_f(z) \rangle = 0$ .  $\square$

**Remark A.1.** For  $n = 2$ , the proof of Theorem A.1 shows that  $\nabla d_f(z)$  is orthogonal to  $[e^{i\psi_l}, e^{i\psi_k}]$  and may vanish for certain  $z$ . In fact, Equation (8) implies that the harmonic measure of the arc  $(\theta_1, \theta_2)$  at any  $z$  on the circle through  $e^{i\theta_1}, e^{i\theta_2}$ , and the origin is equal to the harmonic measure at the origin itself, so that  $d_f(z) = 0$  and  $\nabla d_f(z) = 0$ .