

On the Smallest Enclosing Information Disk*

Frank Nielsen[†]

Richard Nock[‡]

Abstract

In this paper, we present a generalization of the smallest enclosing disk problem for point sets lying in Information-geometric spaces. Given a set of vector points equipped with a (dis)similarity measure that is not necessarily the Euclidean distance, we investigate the problem of finding its center defined as the point minimizing the maximum distance to the point set. For a broad class of distortion measures known as Bregman divergences, these centers are unique and can further be computed efficiently in linear-time by extending the randomized Euclidean algorithm of Welzl [7]. As an application, we show how to solve a statistical model estimation problem by computing the center of a finite set of 1D Normal distributions.

1 Introduction

Given a set $\mathcal{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ of n vector points, we are interested in computing a simplified description, which is a good fit to \mathcal{S} , called its *center* \mathbf{c}^* . Two optimization criteria are usually considered for finding \mathbf{c}^* : (MINAVG) minimizes the *average distortion* $\mathbf{c}^* = \operatorname{argmin}_{\mathbf{c}} \sum_i d(\mathbf{c}, \mathbf{s}_i)$, or (MINMAX) minimizes the *maximal distortion* $\mathbf{c}^* = \operatorname{argmin}_{\mathbf{c}} \max_i d(\mathbf{c}, \mathbf{s}_i)$. These problems have been widely studied in computational geometry (1-center problem), computational statistics (1-point estimator), and machine learning (1-class classification). It is known that for *squared* Euclidean distance (L_2^2) the *centroid* is the center of MINAVG(L_2^2) [2]. For the Euclidean distance L_2 , the circumcenter of \mathcal{S} is the center of MINMAX(L_2), and the Fermat-Weber point is the center of MINAVG(L_2). Historically, the smallest enclosing disk problem dates back to 1857 when Sylvester first asked for the smallest disk enclosing n points on the Euclidean plane \mathbb{E}^2 , a metric space. Although $O(n \log n)$ -time algorithms were designed for the planar case in the early 1970s, the problem complexity was only settled in 1984 with Megiddo’s first linear time algorithm for solving linear programs. Unfortunately, Megiddo’s pruning algorithm exhibits a large constant

hidden in the big-Oh notation and performs poorly in practice. Welzl [7] developed a simple and elegant recursive $\tilde{O}(n)$ randomized algorithm that is often used in practice. On the Euclidean plane \mathbb{E}^2 , the distance measure $d(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|$ defines a metric space (the L_2 norm). In a metric space, the distance function has important properties: (1) $d(\mathbf{p}, \mathbf{q}) \geq 0$ with equality if and only if $\mathbf{p} = \mathbf{q}$, (2) symmetry $d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p})$, and (3) triangle inequality: $d(\mathbf{p}, \mathbf{r}) \leq d(\mathbf{p}, \mathbf{q}) + d(\mathbf{q}, \mathbf{r})$. A disk $\mathcal{B} = \text{Disk}(\mathbf{c}, r)$ of center \mathbf{c} and radius r is defined as the set of points that are within distance r from the center: $\mathcal{B} = \{\mathbf{x} \in \mathbb{E}^2 \mid d(\mathbf{c}, \mathbf{x}) \leq r\}$. Euclidean disks are invariant by translation: $\mathcal{B} + \mathbf{t} = \text{Disk}(\mathbf{c} + \mathbf{t}, r)$.

In computational machine learning, it is seldomly the case that the L_2 geometric distance reflects the distortion between two d -dimensional data elements (vector points are information containers). A general distortion framework, known as *Bregman divergences* [2, 6], is rather used. Informally speaking, a Bregman divergence D_F is the tail of a Taylor expansion of a *strictly convex and differentiable* function F :

$D_F(\mathbf{x}, \mathbf{y}) = F(\mathbf{x}) - F(\mathbf{y}) - \langle \mathbf{x} - \mathbf{y}, \nabla F(\mathbf{y}) \rangle$, where ∇F denotes the gradient operator, and $\langle \cdot, \cdot \rangle$ the inner product (dot product). Bregman divergences are parameterized families of distortions defined on a convex domain $\mathcal{X} \subseteq \mathbb{R}^d$ for strictly convex and differentiable functions F on $\text{int}(\mathcal{X})$ (see Figure 1 and Table 1). For Bregman divergences, there exist two types of Bregman balls depending on the argument position of the center [6]: $\mathcal{B}_{\mathbf{c}, r} = \{\mathbf{x} \in \mathcal{X} : D_F(\mathbf{c}, \mathbf{x}) \leq r\}$ and $\mathcal{B}'_{\mathbf{c}, r} = \{\mathbf{x} \in \mathcal{X} : D_F(\mathbf{x}, \mathbf{c}) \leq r\}$, that are not necessarily convex nor identical. We can further define a third-type of disk by taking the symmetric divergence

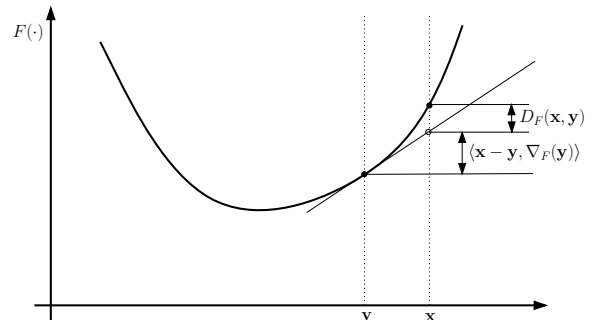


Figure 1: Visualizing convex and differentiable function F and its corresponding Bregman divergence $D_F(\cdot, \cdot)$.

*Applet available at <http://www.cs1.sony.co.jp/person/nielsen/BregmanBall/MINIBALL/>

[†]Fundamental Research Laboratory, Sony Computer Science Laboratories Inc., Japan, Frank.Nielsen@acm.org

[‡]DSI-GRIMAAG, University of Antille-Guyane, Martinique, France rnock@martinique.univ-ag.fr

$D'_F(\mathbf{x}, \mathbf{c}) = D'_F(\mathbf{c}, \mathbf{x}) = \frac{D_F(\mathbf{x}, \mathbf{c}) + D_F(\mathbf{c}, \mathbf{x})}{2}$. Note that Bregman disks are not invariant by translation. For the remainder of this paper, we consider only the first-type of disks $\mathcal{B}_{\mathbf{c}, r}$.

We have shown in [6] that Bregman smallest enclosing balls are unique, thus generalizing the former results of Welzl for balls/ellipsoids [7]. We denote by $\mathcal{B}^* \mathcal{S}$ the smallest enclosing ball of set \mathcal{S} . Moreover, let $c^*(\mathcal{S})$ and $r^*(\mathcal{S})$ denote the center and radius of the smallest enclosing ball $\mathcal{B}^* \mathcal{S}$ of \mathcal{S} .

Domain	$F(\mathbf{s})$	$D_F(\mathbf{c}, \mathbf{s})$
	Quadratic loss	L_2^2 norm
\mathbb{R}^d	$\sum_{j=1}^d s_j^2$	$\sum_{j=1}^d (c_j - s_j)^2$
$(\mathbb{R}^{+,*})^d$	Shannon entropy	(I/KL)-divergence
d -simplex	$\sum_{j=1}^d s_j \log s_j - s_j$	$\sum_{j=1}^d c_j \log(c_j/s_j) - c_j + s_j$
$(\mathbb{R}^{+,*})^d$	Burg entropy	Itakura-Saito divergence
	$-\sum_{j=1}^d \log s_j$	$\sum_{j=1}^d (c_j/s_j) - \log(c_j/s_j) - 1$

Table 1: Most common Bregman divergences D_F .

2 LP-type and Basis Procedures

The randomized linear-time algorithm of Welzl [7] for finding the smallest enclosing ellipsoid is a particular case of a broader class of algorithms that solve linear programming-type (LP-type) problems [3]. Finding the smallest Bregman ball is LP-type because it satisfies the two sufficient and necessary LP-type axioms [3]:

Monotonicity. For any \mathcal{F} and \mathcal{G} such that $\mathcal{F} \subseteq \mathcal{G} \subseteq \mathcal{X}$, $r^*(\mathcal{F}) \leq r^*(\mathcal{G})$.

Locality. For any \mathcal{F} and \mathcal{G} such that $\mathcal{F} \subseteq \mathcal{G} \subseteq \mathcal{X}$ with $r^*(\mathcal{F}) = r^*(\mathcal{G})$, and any point $\mathbf{p} \in \mathcal{X}$,

$$r^*(\mathcal{G}) < r^*(\mathcal{G} \cup \{\mathbf{p}\}) \rightarrow r^*(\mathcal{F}) < r^*(\mathcal{F} \cup \{\mathbf{p}\}).$$

The latter locality property holds because of the uniqueness of Bregman balls. Thus, we are able to use Welzl's abstract randomized recursive algorithm [7]:

```

MINIINFOBALL( $\mathcal{S} = \{\mathbf{p}_1, \dots, \mathbf{p}_n\}, \mathcal{B}$ )
1.  $\triangleleft$  Initially  $\mathcal{B} = \emptyset$ . Returns  $B^* = (c^*, r^*) \triangleright$ 
2. if  $|\mathcal{S} \cup \mathcal{B}| \leq 3$ 
3.   then return  $B = \text{SOLVEINFOBASIS}(\mathcal{S} \cup \mathcal{B})$ 
4.   else
5.     Select at random  $\mathbf{p} \in \mathcal{S}$ 
6.      $B^* = \text{MINIINFOBALL}(\mathcal{S} \setminus \{\mathbf{p}\}, \mathcal{B})$ 
7.     if  $\mathbf{p} \notin B^*$ 
8.       then  $\triangleleft$  Then add  $\mathbf{p}$  to the basis  $\triangleright$ 
9.       return  $\text{MINIINFOBALL}(\mathcal{S} \setminus \{\mathbf{p}\}, \mathcal{B} \cup \{\mathbf{p}\})$ 
    
```

We still need to solve the basis problem: solving the smallest enclosing disk of (at most) three points \mathcal{B} . We do this by computing all enclosing disks of \mathcal{B} generated by either two or three points of \mathcal{B} on their boundaries,

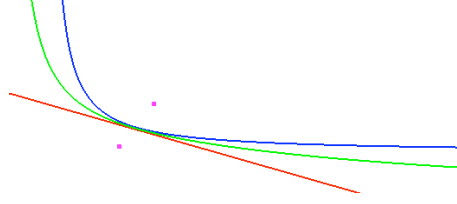


Figure 2: Three Itakura-Saito bisectors: first-type (red), second-type (blue) and third-type (green). The first-type Bregman bisector is *always* a linear separator.

and choose the disk that has minimum radius (*ie.*, minimum divergence). For computing *exactly* the center of a Bregman disk passing through three points, we first define the Bregman bisectors. Let $\text{Bisector}(\mathbf{p}, \mathbf{q}) = \{\mathbf{c} \in \mathcal{X} \mid D_F(\mathbf{c}, \mathbf{p}) = D_F(\mathbf{c}, \mathbf{q})\}$ be the Bregman bisector of locii \mathbf{p} and \mathbf{q} . That is, $\text{Bisector}(\mathbf{p}, \mathbf{q})$ represents the set of points that have the same divergence to \mathbf{p} and \mathbf{q} . We observe the following lemma:

Lemma 1 *The first-type Bregman bisector is linear. (But not necessarily the second- nor the third-type, as depicted in Figure 2.)*

Proof. We write $D_F(\mathbf{c}, \mathbf{p}) = D_F(\mathbf{c}, \mathbf{q})$. That is,

$$\begin{aligned}
 F(\mathbf{c}) - F(\mathbf{p}) - \langle (\mathbf{c} - \mathbf{p}), \nabla F(\mathbf{p}) \rangle &= \\
 F(\mathbf{c}) - F(\mathbf{q}) - \langle (\mathbf{c} - \mathbf{q}), \nabla F(\mathbf{q}) \rangle, & \\
 \langle \mathbf{c}, (\nabla F(\mathbf{p}) - \nabla F(\mathbf{q})) \rangle + & \\
 F(\mathbf{p}) - F(\mathbf{q}) + \langle \mathbf{q}, \nabla F(\mathbf{q}) \rangle - \langle \mathbf{p}, \nabla F(\mathbf{p}) \rangle &= 0.
 \end{aligned}$$

This is a linear equation in \mathbf{c} . Thus, the bisector $\text{Bisector}(\mathbf{p}, \mathbf{q}) = \{\mathbf{x} \mid \langle \mathbf{x}, \mathbf{d}_{\mathbf{p}\mathbf{q}} \rangle + k_{\mathbf{p}\mathbf{q}} = 0\}$ (with $\mathbf{d}_{\mathbf{p}\mathbf{q}} = \nabla F(\mathbf{p}) - \nabla F(\mathbf{q})$ a vector and $k_{\mathbf{p}\mathbf{q}} = F(\mathbf{p}) - F(\mathbf{q}) + \langle \mathbf{q}, \nabla F(\mathbf{q}) \rangle - \langle \mathbf{p}, \nabla F(\mathbf{p}) \rangle$ a constant) is geometrically an hyperplane (*eg.*, a line for 2D vectors). \square

It follows that for *any* Bregman divergence, the exact circumcenter of the Bregman disk passing through three points $\mathbf{p}_1, \mathbf{p}_2$ and \mathbf{p}_3 can be computed *exactly* as the intersection point of any two bisectors. We get: $\mathbf{c}^* = l_{12} \times l_{13} = l_{12} \times l_{23} = l_{13} \times l_{23}$, where l_{ij} is the projective line associated to the linear bisector $\text{Bisector}(\mathbf{p}_i, \mathbf{p}_j)$ and \times denote the vector cross-product operation (see [4] for further explanations/source codes). That is, the ‘‘circumcenter’’ of three points is the Bregman *trisector*, as shown in Figure 3. Observe that although we compute the circumcenter exactly, the border of the Bregman ball is rasterized approximately (require to solve a convex optimization). To solve for the exact circumcenter \mathbf{c}^* of the smallest Bregman disk passing through two points \mathbf{p} and \mathbf{q} , we consider $\mathbf{c}^* \in \text{Bisector}(\mathbf{p}, \mathbf{q})$ and minimize $D_F(\mathbf{c}, \mathbf{p})$. Another approach is to perform a logarithmic search on parameter $\lambda \in [0, 1]$ used to describe a point $\mathbf{r}_\lambda = \nabla F^{-1}((1 - \lambda)\nabla F(\mathbf{p}) + \lambda\nabla F(\mathbf{q}))$

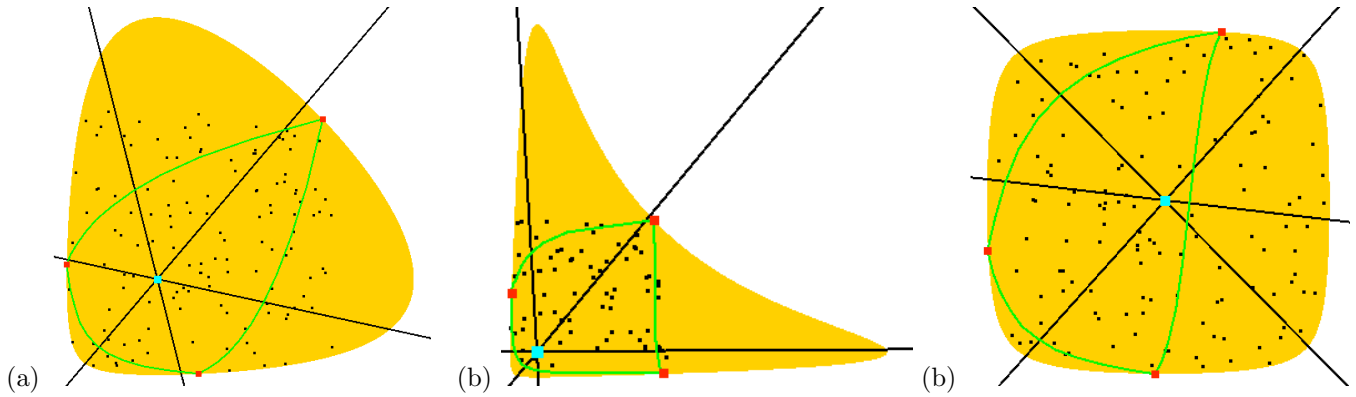


Figure 3: Examples of smallest enclosing disks for various Bregman divergences: (a) Kullback-Leibler, (b) Itakura-Saito and (c) Logistic loss. See applet at <http://www.csl.sony.co.jp/person/nielsen/BregmanBall/MINIBALL/>

on the geodesic of $\mathbf{p}\mathbf{q}$, where ∇_F^{-1} denotes the reciprocal gradient. For Mahalanobis distance (including squared Euclidean distance), \mathbf{c}^* is simply the intersection of the bisector with the line passing through \mathbf{p} and \mathbf{q} (for L_2 , it is simply $\frac{\mathbf{p}+\mathbf{q}}{2}$). This is not always the case for Bregman divergences (eg., Kullback-Leibler or Itakura-Saito divergences). However, for *any* Bregman divergence, we get a convex optimization problem that can be solved approximately within b bits (machine precision) in $O(b)$ time. In fact, we do not need to compute the circumcenter of the disk, but rather implicitly represent and manipulate the disk using its combinatorial basis (two or three points). To decide whether a point \mathbf{p} falls inside, on, or outside the Bregman disk defined by two points \mathbf{p}_1 and \mathbf{p}_2 , We compute the exact radius of the disk $\text{Disk}(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p})$ and check whether that radius is strictly larger than the radius of the smallest disk $\text{Disk}(\mathbf{p}_1, \mathbf{p}_2)$ or not. This becomes a *decision problem* (better known as the *InSphere* predicate) that can be solved only within some prescribed precision (bit complexity model). (Note that Bregman co-circularities can be detected exactly by checking the centers of any 3-point subsets.)

3 An Application Example of MINIINFOBALL

Let \mathcal{N} denote the statistical exponential family of univariate Normal distributions.¹ A Normal probability distribution $N(\mu, \sigma) \in \mathcal{N}$ with mean μ and variance σ^2 (σ is called the standard deviation) has a corresponding probability density function² $N(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$ (with $\forall x N(x|\mu, \sigma) \geq 0$ and $\int_{-\infty}^{\infty} N(x|\mu, \sigma)dx = 1$). Let $\mathcal{S} = \{N_1, \dots, N_n\}$ be a set of n univariate Normal distributions $N_i = N(\mu_i, \sigma_i)$. Estimating the population center of \mathcal{S} amounts to finding the Normal distribution $N^*(\mu^*, \sigma^*) \in \mathcal{N}$ such that the

maximum divergence of N^* to any $N_i \in \mathcal{S}$ is minimized. That is, the population mean μ^* and population variance σ^{*2} defines the center of the smallest enclosing disk of the 2D set $\mathcal{P} = \{\mathbf{p}_1, \dots, \mathbf{p}_n\}$ with $\mathbf{p}_i = (\mu_i, \sigma_i)$ for all $i \in \{1, \dots, n\}$. This statistical model selection is useful for person/machine identification or change detection algorithms that require to take into account variability of distributions, as depicted in Figure 5. We need to choose an appropriate divergence D_F for Normal distributions. The entropy $H(f)$ of a pdf f is defined as $H(f) = \int_x f(x) \log_2 \frac{1}{f(x)}$, and mathematically represents the amount of information in bits.³ The relative entropy KL, also known as the Kullback-Leibler divergence [1], measures the dissimilarity of two probability distributions f and g : $\text{KL}(f, g) = \int_x f(x) \log_2 \frac{f(x)}{g(x)}$. The relative entropy corresponds to the average number of additional bits required for coding f when using an optimal code for g . The measure is therefore not symmetric nor does the triangle inequality hold. Let $H(f, g)$ denote the cross-entropy $-\int_x f(x) \log g(x)$. Then, we have $\text{KL}(f, g) = H(f, g) - H(f)$ (entropy loss). It turns out that the relative entropy KL is a good distortion measure of distributions, and belongs to the family of Bregman divergences. In fact, we can rewrite the pdf of the Normal distribution as $N(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}Z(\boldsymbol{\theta})} \exp\{\langle \boldsymbol{\theta}, \mathbf{f}(x) \rangle\}$, with $Z(\boldsymbol{\theta}) = \sigma \exp\{\frac{\mu^2}{2\sigma^2}\} = \sqrt{-\frac{1}{2\theta_1} \exp\{-\frac{\theta_2^2}{4\theta_1}\}}$, $\mathbf{f}(x) = [x^2 \ x]^T$ is called the *sufficient statistics* and $\boldsymbol{\theta} = [-\frac{1}{2\sigma^2} \ \frac{\mu}{\sigma^2}]^T$ is called the *natural parameters* of the *statistical exponential family* of Normal distributions. Exponential families contain many famous distributions such as Poisson, Gaussian and multinomial distributions, and have been thoroughly

³For example, assume that f is the frequency of letter occurrences in a given text, then $H(f) = \sum_x f(x) \log_2 \frac{1}{f(x)}$ provides a theoretical lower bound on the minimum number of bits required to encode each letter (see Huffman or Ziv-Lempel codings). For natural logarithm (base e), $H(f) = \int_x f(x) \log \frac{1}{f(x)}$, the unit of information of the Shannon entropy is defined as a *nat*.

¹Also commonly called Gaussian distributions.

²pdf for short, the relative frequency function.

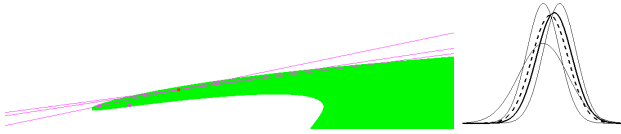


Figure 4: “Normal” disk passing through exactly three Normal points in the coordinate system $(\mu^2 + \sigma^2, \mu)$. Thick Gaussian corresponds to the MINMAX center. Thick dashed Gaussian depicts the MINAVG center.

studied in Information Geometry [1]. The Kullback-Leibler divergence between any two models \mathbf{p} and \mathbf{q} of an exponential family with respective parameters θ_p and θ_q is obtained from the Bregman divergence by choosing $F(\theta) = -\log Z(\theta)$. This yields $KL(\theta_p || \theta_q) = D_F(\theta_p, \theta_q) = \langle (\theta_p - \theta_q), \theta_p[\mathbf{f}] \rangle + \log \frac{Z(\theta_q)}{Z(\theta_p)}$, with:

$$\theta_p[\mathbf{f}] = \begin{bmatrix} \int_x \frac{x^2}{Z(\theta_p)} \exp\{\langle \theta_p, \mathbf{f}(x) \rangle\} \\ \int_x \frac{x}{Z(\theta_p)} \exp\{\langle \theta_p, \mathbf{f}(x) \rangle\} \end{bmatrix} = \begin{bmatrix} \mu_p^2 + \sigma_p^2 \\ \mu_p \end{bmatrix}$$

Note that the equation of the linear bisector is: $\langle (\theta_p - \theta_q), \theta_c[\mathbf{f}] \rangle + \log \frac{Z(\theta_p)}{Z(\theta_q)} = 0$. Thus, we choose to change variables as $(\mu, \sigma) \rightarrow (\mu^2 + \sigma^2, \mu) = (x, y)$ (with $x > y > 0$) to get the proper coordinate system on which to apply MININFOBALL. It comes that $Z(x, y) = \sqrt{x - y^2} \exp\{\frac{y^2}{2(x - y^2)}\}$, $\log Z(x, y) = \log \sqrt{x - y^2} + \frac{y^2}{2(x - y^2)}$ and $\nabla_F(x, y) = (\frac{1}{2(x - y^2)} - \frac{y^2}{2(x - y^2)^2}, \frac{y^3}{(x - y^2)^2})$. Once the center (x^*, y^*) of the smallest enclosing ball is computed, we retrieve the corresponding (μ^*, σ^*) parameters as $(y^*, x^* - (y^*)^2)$. Figure 4 depicts the ball passing through exactly three Normal “statistical” points $(\mu_1 = 2, \sigma_1 = 1)$, $(\mu_2 = 2, \sigma_2 = \frac{3}{2})$ and $(\mu_3 = 3, \sigma_3 = 1)$. We found $(\mu^*, \sigma^*) \simeq (2.67446, 1.08313)$ and $r^* \simeq 0.801357$. For comparison, the MINAVG center has coordinates $(\mu^*, \sigma^*) = (2.40909, 1.10782)$. Computing the reciprocal gradient $\nabla_F^{-1}(x, y)$ requires to solve a quartic equation, that we omit here. Note that for two normal distributions $N_i(\mu_i, \sigma_i)$ and $N_j(\mu_j, \sigma_j)$, the relative Kullback-Leibler entropy $KL(N_i, N_j)$ admits the following closed-form solution: $KL(N_i, N_j) = \frac{1}{2} \left(\frac{\sigma_j^2}{\sigma_i^2} + 2 \log \frac{\sigma_j}{\sigma_i} - 1 + \frac{(\mu_j - \mu_i)^2}{\sigma_j^2} \right)$. The mean and variance parameters of the Gaussian distributions can be estimated from random samples using the maximum likelihood estimators (MLEs). The MLEs of μ and σ^2 given by a random sample $\mathbf{n} = \{n_1, \dots, n_k\}$ from $N(x|\mu, \sigma)$ is $\hat{\mu} = \frac{1}{k} \sum_{i=1}^k n_i$ and $\hat{\sigma} = \frac{1}{k} \sum_{i=1}^k (n_i - \hat{\mu})^2$, respectively. If we assume the standard deviations identical, this KL-divergence becomes simply a weighted squared Euclidean distance. In sound processing, since it has been observed that human ear perceives loudness in a logarithmic way, the Itakura-Saito divergence (see Table 1) often measures the perceptual closeness of two power spectra. We can similarly ask for the center spec-

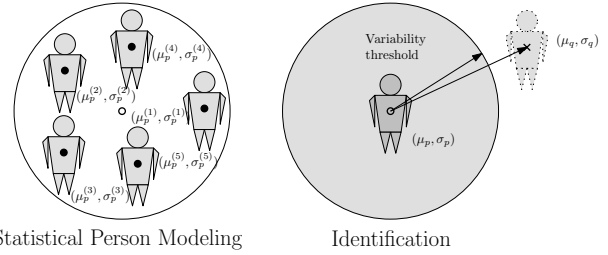


Figure 5: Identification/change detection by model selection: each observation of a person P yields a statistical Normal distribution $(\mu_p^{(i)}, \sigma_p^{(i)})$. A person is modeled using a Normal distribution (μ_p, σ_p) by minimizing the MinMax KL divergence and a variability parameter r : the radius of the smallest enclosing Bregman disk. A person Q is different from P iff $KL(P||Q) > r$.

trum that minimizes the maximal divergence to a given set of spectra, a problem known in the statistics literature as model/feature selection. This again amounts to compute the smallest enclosing Itakura-Saito (Bregman) ball.

4 Perspectives

The methodology extends to arbitrary dimension as well but requires to solve convex optimization programs for finding out centers of basis. For large dimensions, we have shown how to extend the simple 2-line iterative approximation algorithm of Bădoiu and Clarkson to arbitrary Bregman divergences [6] (based on the existence of core-sets). Computing exactly the center of the information ball passing through $d + 1$ points is a key primitive to generalize the Voronoi/Delaunay structures on information-theoretic spaces of dimension d using the notion of “empty spheres.”

References

- [1] S.-I. Amari and H. Nagaoka. *Methods of Information Geometry*, Vol. 191, AMS, 2000.
- [2] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh. Clustering with Bregman divergences. *Journal of Machine Learning Research*, 6:1705–1749, 2005.
- [3] J. Matoušek, M. Sharir, and E. Welzl. A subexponential bound for linear programming. 16(4–5):498–516, 1996.
- [4] F. Nielsen. *Visual Computing: Geometry, Graphics, and Vision*. Charles River Media, ISBN 1584504277, 2005.
- [5] F. Nielsen and R. Nock. Approximating smallest enclosing balls. In *ICCSA CGA*, LNCS #3045. Springer, 2004.
- [6] R. Nock and F. Nielsen. Fitting the smallest enclosing Bregman ball. In *ECML*, LNCS #3720, 2005.
- [7] E. Welzl. Smallest enclosing disks (balls and ellipsoids). *Lecture Notes Comput. Sci.*, 555:359–370, 1991.