Discrepancy norm: approximation and variations

Jean-Luc Bouchot*and Frank Bauer†

May 23, 2014

Abstract

This paper introduces an approach for the minimization of the discrepancy norm. The general idea is to replace the infinity norms appearing in the definition by L^p norms which are differentiable and to make use of this approximation for local optimization.

We will show that the discrepancy norm can be approximated up to any ε and the robustness of this approximation is shown. Moreover, analytical formulation of the derivative of the discrepancy correlation function is given.

In a following step we extend the results to higher dimensional data and derive the related forms for approximations and differentiations.

1 Introduction

This paper deals with the use of the discrepancy norm as introduced by Weyl [1] in the context of signal analysis. A great deal of research in discrepancy theory has been done into numerical integration [2], analysis of randomness [3] and uniformity [4], low-discrepancy sequences for sampling [5], computational geometry [6], and geometric approximation [7]. We invite the reader to have a look at the book of Chazelle for a rather complete review of topics in discrepancy theory [8].

Here we investigate another application of discrepancy theory, namely to signal analysis. As re-introduced by Neunzert and Welton [9] and later by Moser [10] the discrepancy norm becomes an interesting tool for pattern recognition. We revisit here some previous work [11] where an approximation of the discrepancy norm was developed for discrete times series analysis. We extend our results to continuous measurable functions and compute its analytic derivative that can be used for optimization purposes.

In a second step we generalize these results to higher-dimensional signals by considering directional discrepancies and combining them together. Artificial experiments are provided to validate our theoretical findings.

Our contribution is two-fold. We first extend previous results [11] obtained for one-dimensional discrete signals to the more general case of continuous measurable functions. We show that our approximation can be done up to any ε for any function from a measurable space with finite measure. These results are also proven in higher dimension. Second we compute the derivative of this approximation that comes handy for optimization (for instance in the context of signal alignment).

We first start by reviewing the basics regarding the discrepancy norm in Sec. 2: its definition and its useful properties. Of particular interest for signal alignment is the monotonicity property of its autocorrelation function (Property C3). Then Sec. 3 introduces the approximation through L^p-norms and its derivative for one dimensional data. We show some convergence criteria to prove the robustness of the approaches. Sec. 4 extends the previous results to higher dimensional data. We see how the problem arise in the discrepancy norm's definition when dealing with higher dimensions. Numerical experiments are given along the theoretical results.

*Department of Mathematics, Drexel University, 3141 Chstenut Street, Philadelphia

†DZ BANK AG, Kapitalmärkte Handel, Frankfurt am Main, Germany

2 Discrepancy norm

2.1 Introduction and first definitions

Discrepancy measurements have been studied for a very long time and date back to Hermann Weyl's theory [1]. Here we are more interested in its application in terms of similarity measure [12].

As seen in [13, 14], the discrepancy norm introduced here is particularly suited for image registration. Other approaches have been studied recently for pattern recognition [9] and vision purposes [15]. It has been reintroduced for general measurable functions by Moser [10].

Definition 1 (Discrepancy norm). *Let $(\mathbb{R}, \Sigma, \mu)$ be a measure space with μ a finite measure. The discrepancy norm of a function $f \in L(\mathbb{R}, \mu)$ is defined as*

$$\|f\|_D = \sup_{[a,b] \subset \mathbb{R}} \left| \int_a^b f \mathrm{d}\mu \right|$$

2.2 Properties

It has been proven that the discrepancy norm shows particularly interesting properties for analyzing misalignments in signals [12] compared to other traditional measures.

Definition 2 (Misalignment functions). *Given a distance measure*

$$d : L(\mathbb{R}, \mu) \times L(\mathbb{R}, \mu) \to \mathbb{R}^+$$

one can define a misalignment measure of a signal as follows:

$$\Delta_d[f](t) := d(f, f_t)$$

where f_t is a translation of f defined as $\forall x \in \mathbb{R}, f_t(x) = f(x - t)$.

In [10] three criteria have been introduced to evaluate the ability of a metric to act as a misalignment measure:

C1 Positive-definiteness:

$$\Delta_d[f](t) = 0 \Leftrightarrow t = 0$$

C2 Continuity:

$$\Delta_d[f](t) \to 0, \text{ for } t \to 0$$

C3 Monotonicity:

$$\forall 1 \leq \lambda, \forall t \in \mathbb{R}, \ \Delta_d[f](t) \leq \Delta_d[f](\lambda t)$$

Property 1 (Compatibility of the discrepancy norm). *Given any measurable function $f \in L(\mathbb{R}, \mu)$, $f \geq 0, \mu$ almost everywhere, the discrepancy norm fulfills the three criteria described above.*

Note that this result is stated on the real line but it could be adapted on the torus to deal with periodic functions. With such function, criteria C1 and C3 need to be adapted. We actually have a more stronger result regarding the continuity criterion:

Property 2 (Lipschitz continuity). *The variations of the discrepancy misalignment function are bounded by its infinity norm:*

$$\Delta_{\|\cdot\|_D}[f](t) \leq |t| \|f\|_\infty$$

The next property defines another computation for the discrepancy norm which appears to be much faster in the applications

Property 3 (Linear computation). *Let $f \in L(\mathbb{R}, \mu)$, the following holds true*

$$\|f\|_D = \max_{b \in \mathbb{R}} \int_{-\infty}^b f \mathrm{d}\mu - \min_{a \in \mathbb{R}} \int_{-\infty}^a f \mathrm{d}\mu$$

This last formulation allows us to compute the discrepancy of a function by means of integral images [16, 17]. These integral images are, in practice, computed in a linear time with respect to the number of samples.

3 Approximation and optimization: the one dimensional case

In this section we extend the results shown in [11] in the context of discrete one dimensional signals. We extend here this approach to more general measurable functions and derive formula to assess the quality of the approximation. Finally a formula for the derivative of the discrepancy correlation function is given in the last subsection.

3.1 Approximation of discrete signals

This subsection is intended as an introduction for our next results and is based on the work of Bouchot et al. [11].

In this case, we consider the counting measure μ_c and function of bounded support. Without loss of generality, we have $f : D := \{1, \cdots, N\} \to \mathbb{R}$ and the discrepancy norm reads

$$\|f\|_D = \max_{i,j \in D} \left| \sum_{k=i}^{j} f_k \right| = \max_{j \in D} \left\{ 0, \sum_{k=1}^{j} f_k \right\} - \min_{i \in D} \left\{ 0, \sum_{k=1}^{i} f_k \right\}$$

Note that the 0 in the max, min operators is essential as the sums should be taken starting at $-\infty$. But the function can be extended with 0 outside the domain D. We can also equivalently write $\|f\|_D = \max_{j \in D_0} \sum_{k=0}^{j} f_k - \min_{i \in D_0} \sum_{k=0}^{i} f_k$ where we define $D_0 = D \cup \{0\}$ and $f_0 = 0$.

Before we recall the idea of the approximation and its property, we give some definitions

Definition 3 (Generalized mean)**.** *Let ϕ be a continuous invertible function, let $f = \{f_k\}_{k=1}^N$, we can define [18] a mean as*

$$M_\phi(f) = \phi^{-1} \left(\frac{\sum_{k=1}^N \phi(f_k)}{N} \right)$$

Moreover, as we would expect from a mean, we have $\min_k f_k \le M_\phi(f) \le \max_k f_k$. (see [19, Ch II] for the proof)

Now if we consider taking L^p norms as ϕ in the previous definition, this yields the following p means:

$$M_p(f) = \left(\frac{\sum_{k=1}^N |f_k|^p}{N} \right)^{1/p}$$

In our previous work we have introduced an approximation of the discrepancy norm for discrete one dimensional signals as described in the following theorem.

Theorem 1. *Let $f \in \ell^1 = L(\mathbb{Z}, \mu_c)$, we define*

$$\gamma_p : \ell^1 \to \mathbb{R}^+$$
$$f \mapsto \gamma_p(f) := \ln \left(\frac{M_p(\chi(f))}{M_{-p}(\chi(f))} \right)$$

where $\chi(f)(k) := \exp(\sum_{l=0}^k f_l)$ is the exponential of the cumulative function.

The following holds:

$$\gamma_p(f) \le ||f||_D < \gamma_p(f) + \frac{2}{p} \ln(N+1)$$

and γ_p is positive definite.

Proofs and details about this theorem can be found in [11].

Corollary 1 (Choice of p)**.**

$$\forall \varepsilon > 0, \forall f \in \ell^1, \forall p \in \mathbb{R}, p \ge p_0 := \frac{2}{\varepsilon} \ln(N+1) \Rightarrow |\gamma_p(f) - \|f\|_D| \le \varepsilon$$

Proof. According to the previous theorem, we have $|\gamma_p(f) - \|f\|_D| < \frac{2}{p}\ln(N+1)$ so the corollary holds true whenever $\frac{2}{p}\ln(N+1) \leq \varepsilon \Leftrightarrow p \geq p_0$ as given in the statement. $\square$

This last part allows to compute an optimal approximation in a practical way. Indeed, due to the power p coming in the approximation process, the danger of getting overflow increases with better approximations.

Conversely computing the approximation of a signal of 1000 samples at order $p = 8$ creates an error no larger than 0.86256, completely independent of the signal's magnitude.

3.2 Approximation of continuous measurable functions

Now that we have seen how practical and close the approximation can be, we want to extend this to continuous measurable functions. From now on we go back to the setting of measure space on $\mathbb{R}$: $(\mathbb{R}, \Sigma, \mu)$ with μ being a finite measure. In the same way, we can introduce the continuous counter-part of the generalized means

$$M_\phi(f) = \phi^{-1}\left(\frac{\int_{\mathbb{R}} \phi(f(x))\mathrm{d}\mu(x)}{\int_{\mathbb{R}} \mathrm{d}\mu}\right)$$

And the following bounding property still holds $\inf_{x\in\mathbb{R}} f(x) \leq M_\phi(f) \leq \sup_{x\in\mathbb{R}} f(x)$ (where inf and sup are understood as their essential definitions: $\operatorname{ess.sup}(f) = \inf\{C \in \mathbb{R} : \mu\{x : f(x) > \sup f\} = 0\}$). We also redefine the p-means M_p in the same way.

The proposed approximation is reintroduced in the same way (with capital letter for the continuous case). Let us first define χ as the exponential of the cumulative function, i.e.:

$$\chi : \left\{ \begin{array}{ccl} L(\mathbb{R},\mu) & \to & L(\mathbb{R},\mu) \\ f & \mapsto & \chi(f) : \chi(f)(x) = e^{\int_{-\infty}^{x} f(t)\mathrm{d}\mu(t)} \end{array}\right.$$

$$\Gamma_p : \left\{ \begin{array}{ccl} L^1(\mathbb{R},\mu) & \to & \mathbb{R}^+ \\ f & \mapsto & \Gamma_p(f) := \ln\left(\frac{M_p(\chi(f))}{M_{-p}(\chi(f))}\right) \end{array}\right.$$

Note that the function f is not necessarily defined on the whole real line but on an interval or union of intervals $I \subset \mathbb{R}$. In this case we can consider its extension $\widetilde{f}$ on the real line as $\widetilde{f}(x) = f(x)$ if $x \in I$ and $\widetilde{f} = 0$ elsewhere. While this extension is not the unique, nor the smartest, option it is sufficient for now.

The following theorem holds:

Theorem 2 (Approximation of the discrepancy norm for 1D continuous functions)**.** Γ_p *defines an approximation of the discrepancy norm in the sense that*

i) $\forall f \in L(\mathbb{R},\mu), \Gamma_p(f) \underset{p\to\infty}{\longrightarrow} \|f\|_D$

ii) $\Gamma_p(f) \leq \|f\|_D$

iii) $\forall \varepsilon > 0, \exists p^* : \forall p \in \mathbb{R}, p \geq p^* \Rightarrow |\Gamma_p(f) - \|f\|_D| \leq \varepsilon$

Proof. Let $f \in L(\mathbb{R},\mu)$, we need to compute $d_p(f) := \|f\|_D - \Gamma_p(f)$. First we rewrite $\|f\|_D$ another way, so that it gets comparable to Γ_p.

$$\begin{aligned} \|f\|_D &= \sup_b \int_{-\infty}^{b} f\mathrm{d}\mu - \inf_a \int_{-\infty}^{a} f\mathrm{d}\mu \\ &= \sup_b \ln\left(\chi(f)(b)\right) + \sup_a \ln\left(\chi(-f)(a)\right) \\ &= \ln\left(\sup_b \chi(f)(b)\right) + \ln\left(\sup_a \chi(-f)(a)\right) = \ln\left(\|\chi(f)\|_\infty \cdot \|\chi(-f)\|_\infty\right) \end{aligned}$$

Plugging this into the expression of d_p, and noting that
$\Gamma_p(f) = \ln\left(\frac{\|\chi(f)\|_p}{\mu(\mathbb{R})^{1/p}} \cdot \frac{\|\chi(-f)\|_p}{\mu(\mathbb{R})^{1/p}}\right)$ we get:

$$d_p(f) = \ln\left(\frac{\|\chi(f)\|_\infty \mu(\mathbb{R})^{1/p}}{\|\chi(f)\|_p} \cdot \frac{\|\chi(-f)\|_\infty \mu(\mathbb{R})^{1/p}}{\|\chi(-f)\|_p}\right)$$

We make use of the following lemma:

Lemma 1 ($p-$norm approximation of the ∞ norm)**.** *Let f be in $L^q \cap L^\infty$ for a certain $q \in \mathbb{R}$. It holds:*

$$\|f\|_p \xrightarrow[p\to\infty]{} \|f\|_\infty$$

A proof of this lemma can be found in [20, Ch.III, Theorem 14F].

If we apply the previous lemma to $\chi(f)$ and $\chi(-f)$, we get that $\frac{\|\chi(f)\|_\infty}{\|\chi(f)\|_p} \xrightarrow[p\to\infty]{} 1$ and $\mu(\mathbb{R})^{1/p} \xrightarrow[p\to\infty]{} 1$. The same holds for $\chi(-f)$ and therefore we have $d_p(f) \xrightarrow[p\to\infty]{} 0$ which proves the first point of the theorem.

For the second point, we make use of the following inequality:

$$\forall f \in L^p(\mathbb{R},\mu) \cap L^q(\mathbb{R},\mu), 1 \leq p \leq q \leq \infty, \|f\|_p \leq \mu(\mathbb{R})^{1/p-1/q}\|f\|_q \tag{1}$$

Thus applying this inequality to $\chi(f)$ and $\chi(-f)$ yields

$$\ln\left(\frac{\mu(\mathbb{R})^{(1/p)}\|\chi(f)\|_\infty}{\|\chi(f)\|_p}\right) \geq 0, \qquad \text{and} \qquad \ln\left(\frac{\mu(\mathbb{R})^{(1/p)}\|\chi(-f)\|_\infty}{\|\chi(-f)\|_p}\right) \geq 0$$

And therefore, we get the second results by summing up both contributions.

To compute a minimum rank of convergence within ε, we make use of the following lemma:

Lemma 2 (Eventual convergence of L^p means towards ∞ norm)**.** *Let $\bar{\varepsilon} > 0$ and $g \in L^q \cap L^\infty$ for a certain $q \in \mathbb{R}$. Let $p_0 = p_0(t,\bar{\varepsilon},g) = \frac{\ln\left(\frac{\mu(E_g(t))}{\mu(\mathbb{R})}\right)}{\ln\left(\frac{1-\bar{\varepsilon}}{1-t}\right)}$, with $E_g(t) = \{x : |g(x)| > \|g\|_\infty(1-t)\}$. It holds*

$$\forall p > \max\{q, p_0\}, \frac{M_p(g)}{\|g\|_\infty} \geq 1 - \bar{\varepsilon}$$

Proof. The convergence is clear due to the previous lemma and inequality (1). Now fix $\bar{\varepsilon} > 0$ and let $\|g\|_\infty > \bar{\varepsilon} \geq t = \alpha\bar{\varepsilon} > 0$ with $0 < \alpha < 1$ and consider the set $E_g(t) = \{x : |g(x)| > \|g\|_\infty(1-t)\}$, it holds:

$$M_p(f) = \left(\frac{\int_{\mathbb{R}} |g|^p \mathrm{d}\mu}{\mu(\mathbb{R})}\right)^{1/p} \geq \left(\frac{\mu(E_g(t))}{\mu(\mathbb{R})}\right)^{1/p} (\|g\|_\infty(1-t))$$

Based on this, we are looking for a lower-bound p_0 of p such that

$$\frac{M_p(g)}{\|g\|_\infty} \geq 1 - \bar{\varepsilon}$$

We have:

$$M_p(g) \geq \left(\frac{\mu(E_g(t))}{\mu(\mathbb{R})}\right)^{1/p} (\|g\|_\infty(1-t)) \geq \|g\|_\infty(1-\bar{\varepsilon})$$

and hence

$$\left(\frac{\mu(E_g(t))}{\mu(\mathbb{R})}\right)^{1/p} \geq \frac{1-\bar{\varepsilon}}{1-t}, \qquad \text{or } p \geq \frac{\ln\left(\frac{\mu(E_g(t))}{\mu(\mathbb{R})}\right)}{\ln\left(\frac{1-\bar{\varepsilon}}{1-t}\right)}$$

Last expression holds for any $\bar{\varepsilon} > t > 0$. For instance, for $t = \bar{\varepsilon}/2$ (or $\alpha = 1/2$)

$$\forall p \geq p_0 = \frac{\ln\left(\frac{\mu(E_g(\bar{\varepsilon}/2))}{\mu(\mathbb{R})}\right)}{\ln\left(2\frac{1-\bar{\varepsilon}}{2-\bar{\varepsilon}}\right)}, M_p(g) \geq \|g\|_\infty(1-\bar{\varepsilon})$$

This finishes the proof of the lemma. □

Now we apply this lemma to both $\chi(f)$ and $\chi(-f)$ and choosing $\bar{\varepsilon} = 1 - e^{-\varepsilon/2}$ and we therefore get two constants $p^+ = p_0(t,\bar{\varepsilon},\chi(f))$ and $p^- = p_0(t,\bar{\varepsilon},\chi(-f))$ and we have $\forall p \geq p^* := \max\{p^+,p^-\}, \Gamma_p(f) \geq \|f\|_D - \varepsilon$.

This finishes the proof of the theorem. □

Remarks This lower bound depends strongly on the function f through the set $E_f(t)$; this is due to the non-equivalence of norms in the continuous case (in the discrete case, the bounds given for the theorem were crisp and uniform).

In the proof we have made use of an upper level set $E_f(t)$ and have proven that we get an upper bound for $t = \varepsilon/2$. It would work for $t = \varepsilon/4$ too and more generally for all $0 < t < \varepsilon < 1$. We have no idea at the moment what would be the influence of this on the estimated p^*.

Finding a nice p^* is a crucial problem for practical applications. Indeed because we are taking the p^{th} power of an exponential, the risk to encounter overflow is high. Therefore we need to ensure a good approximation but keeping the power p as small as possible.

Example of lower bounds For illustration purposes let us see this effect on some toy examples.

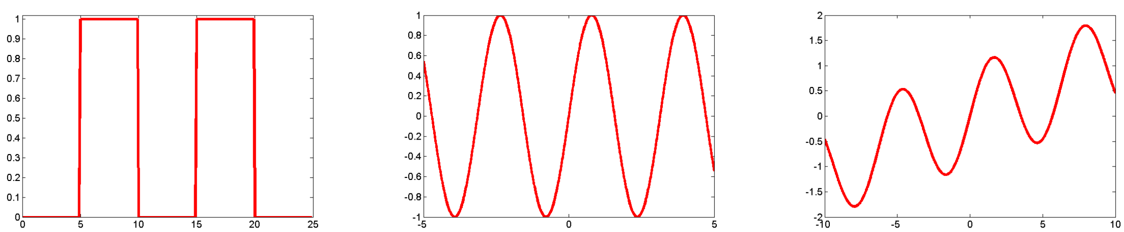
We consider three functions:

$$f_1 : \left\{ \begin{array}{ccl} [0,5] & \rightarrow & \mathbb{R} \\ x & \mapsto & 1 \text{ if } x \in [1,2[\cap[3,4], \ 0 \text{ elsewhere} \end{array} \right.$$

$$f_2 : \left\{ \begin{array}{ccl} [-10,10] & \rightarrow & \mathbb{R} \\ x & \mapsto & \sin(x) \end{array} \right.$$

$$f_3 : \left\{ \begin{array}{ccl} [-10,10] & \rightarrow & \mathbb{R} \\ x & \mapsto & a \cdot x + \sin(x) \end{array} \right.$$

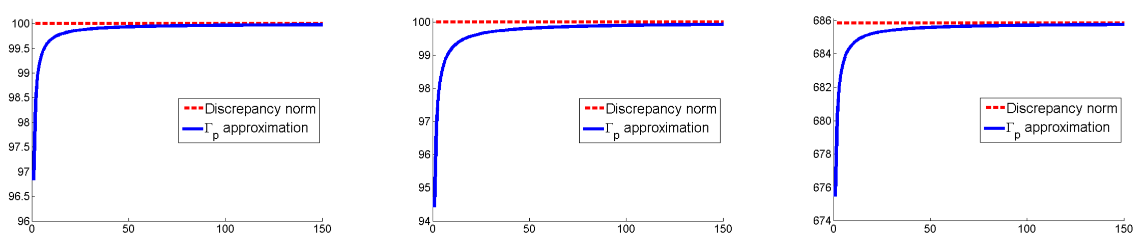
where we have chosen a to be around $1/10$. In our case, we consider a measure as a combination of uniformly distributed diracs (every 0.02 for f_1 and every 0.01 for the two others). f_1, f_2, f_3 are illustrated in Fig. 1.



(a) 2 steps function: 250 data (b) Horizontal wave: 2001 data (c) Oblique wave: 2001 data

Figure 1: Our dataset of 1 dimensional toy functions used in our paper.

We start by analyzing the convergence of the $p-$norms approximations towards the discrepancy norm. The two first statements of Theorem 2 tell that Γ_p gets to $\|\cdot\|_D$ from below with p increasing. Figs. 2 depict this behavior. One sees that the values of Γ_p functions increase drastically with p close to 0 and tends to stabilize to $\|\cdot\|_D$ early.



(a) 2 Steps function (b) Horizontal sine wave (c) Oblique sine wave

Figure 2: Convergence of the Γ_p functions towards the discrepancy norm. The horizontal axis corresponds to changing values of p while the vertical one is the output of the Γ_p and discrepancy norm functions.

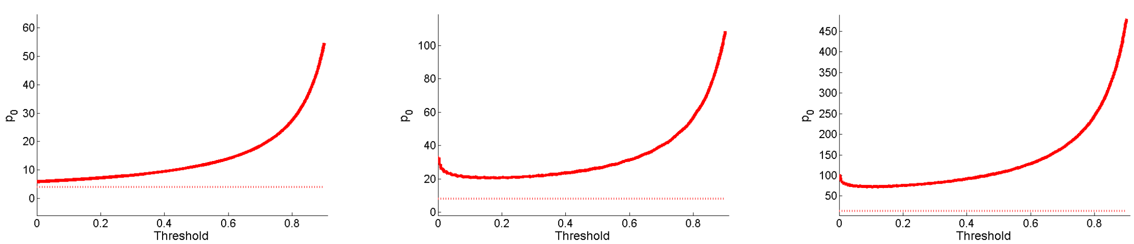
The next point we need to have a look at is the reliability of our estimator for the value of p^*. While this value is clearly not optimal (we have left potentially a lot of the function aside when deriving the estimation), we would like it to be relatively small anyway. As this is the value we are going to use afterwards for our tests, we need it to be small enough, in order not to get into machine overflow.

We have gathered results based on our three toy functions. For all of them, we have noted the p^* defined in Theorem 2, and the smallest integer value, denoted by $\overline{p^*}$, for which the approximation up to an ε_0 holds. Moreover, as it can be seen in the proof of the theorem, one can tune the threshold used for the decomposition of the function. This threshold has also been analyzed for the horizontal sine wave function. The results are summarized in Table. 1.

Function	ε_0	α	p^*	$\Gamma_{p^*}(f)$	$\|f\|_D$	$\overline{p^*}$
2 steps	0.1	1/2	100.77	99.97	100	32
	5	0.2	2.68	98.81	100	1
Horizontal sines	0.1	1/4	257.73	99.96	100.00	101
	0.1	1/2	348.72	99.97	100.00	101
	0.1	3/4	666.56	99.98	100.00	101
	0.05	1/2	767.06	99.98	100.00	212
	2	1/2	13.45	99.39	100.00	4
Oblique sines	1	1/2	105.99	685.72	685.85	13
	2	1/4	44.64	685.55	685.85	7

Table 1: Examples of approximation and lower bound estimations on some toy functions.

It comes out that we indeed get a close approximate of the discrepancy norm of a function by means of $p-$norms approximation. The estimation of a good p^*, however, is strongly dependent on the choice of the threshold. Fortunately first tests tend to show that a threshold in the order of 0.1 or 0.2 yields stable and reliable results.



(a) 2 Steps function (b) Horizontal sine wave (c) Oblique sine wave

Figure 3: Computed values of the p^* estimator. The continuous curves show the estimated p^* while the the dotted straight line corresponds to the $\overline{p^*}$ as described in the Table 1

Figs. 3 show the behavior of the estimation of p^* for different thresholding values. For the rest of the paper we use in our applications a threshold value of 0.2 which seems to lie in a rather stable area and which achieves low values of p^*.

Other properties - Convolution One advantage of this approximation lies in the possibility to solve any misalignment problem. Indeed if we consider f_t as the function f translated by an offset t, we can calculate the correlation function $h(t) = \Gamma_p(f_t - g)$ by simply computing two convolutions. As it was first shown in [11] for the discrete case, we show here that the arguments still hold in case of measurable functions.

Theorem 3. *Let f and g be two measurable functions in the finite σ-algebra, h can be computed by means of two convolution operators (up to a normalization factor):*

$$h = \frac{1}{p} \ln \left(\chi_p(f) * \chi_p(-g) \right) + \frac{1}{p} \ln \left(\chi_p(g) * \chi_p(-f) \right)$$

where $\chi_p(x) = e^{p \int_{-\infty}^{x} f(y) \mathrm{d}\mu(y)}$

Proof. We have

$$h(t) = \Gamma_p(f_t - g) = \frac{1}{p}\ln\left(\int_{\mathbb{R}} \chi_p\left(f_t - g\right)\right) + \frac{1}{p}\ln\left(\int_{\mathbb{R}} \chi_p\left(g - f_t\right)\right) - 2\ln\left(\mu\left(\mathbb{R}\right)\right).$$

Let us consider only the first part of this equation (and the second part is done exactly the same way).

It holds:

$$\begin{aligned}\int_{\mathbb{R}} \chi_p\left(f_t - g\right)(x)\mathrm{d}\mu(x) &= \int_{\mathbb{R}} e^{p\int_{-\infty}^{x-t} f(y)\mathrm{d}\mu(y) - p\int_{-\infty}^{x} g(y)\mathrm{d}\mu(y)}\mathrm{d}\mu(x)\\ &= \int_{\mathbb{R}} \chi_p(f)(x-t)\cdot\chi_p(-g)(x)\mathrm{d}\mu(x)\end{aligned}$$

which is the definition of the correlation of two functions. The second term directly gives:

$$\int_{\mathbb{R}} \chi_p\left(g - f_t\right)(x)\mathrm{d}\mu(x) = \int_{\mathbb{R}} \chi_p(g)(x)\cdot\chi_p(-f)(x-t)\mathrm{d}\mu(x)$$

Hence, $h(t)$ can be computed for all t with only two convolutions. □

This formula can become very handy to speed up the calculations whenever needed.

3.3 Differentiation of the discrepancy misalignment function

In this subsection we want to introduce a way of optimizing the discrepancy norm based correlation function. Given a function f and a reference g, we want to find the optimal translation parameter t^* such that $\|f_{t^*} - g\|_D$ is minimal. We consider the following optimization problem:

$$t^* = \underset{t\in\mathbb{R}}{\operatorname{argmin}}\, \|f_t - g\|_D$$

We would like to optimize the objective function $J(t) = \|f_t - g\|_D$ making use of the monotonicity property of the discrepancy norm when facing misaligned functions. However, due to its definition with max and min, the discrepancy norm is not everywhere differentiable (it is however almost everywhere differentiable due to Rademacher's theorem). Moreover, as it can be seen from Fig. 4, the objective function shows some kind of plateau in some pathological cases which yield any gradient based method to fail in finding a correct solution. Therefore, we want to make use of the previous approximation to compute an approximated gradient to the objective function. As we demonstrate, we can derive a formula based on a scalar product which allows us to compute the derivative very efficiently.

Theorem 4 (Derivative of the discrepancy correlation)**.** *Given a continuously differentiable function f and a function g in $L\left(\mathbb{R};\mu\right)$ the derivative of $J(t)$ can be approximated in the following way:*

$$\frac{\partial J}{\partial t}(t) \approx \left\langle \int_{x=-\infty}^{\cdot} f'(x-t)\,\mathrm{d}\mu(x), \left(\frac{\chi(-f_t+g)}{\|\chi(-f_t+g)\|_p}\right)^p - \left(\frac{\chi(f_t-g)}{\|\chi(f_t-g)\|_p}\right)^p\right\rangle \tag{2}$$

Proof. We need to compute the derivative of the discrepancy norm of a difference. As it has been seen in the previous section, it can be approximated by means of L^p norms and we get

$$\begin{aligned}\frac{\partial J}{\partial t}(t) &\approx \frac{\partial \Gamma_p}{\partial t}(f_t - g) = \frac{\partial}{\partial t}\left(\frac{1}{p}\ln\left(\|\chi(f_t-g)\|_p^p\right) + \frac{1}{p}\ln\left(\|\chi(-f_t+g)\|_p^p\right) + C\right)\\ &= \frac{1}{p}\frac{\partial}{\partial t}\left(\Gamma_p^{(+)}(f_t-g) + \Gamma_p^{(-)}(f_t-g)\right)\end{aligned}$$

Where we have defined $\Gamma_p^{(+)}(h) := \ln\left(\|\chi(h)\|_p^p\right) := \Gamma_p^{(-)}(-h)$,

$$\frac{\partial \Gamma_p^{(\pm)}}{\partial t}(f_t - g) = \frac{1}{\|\chi(\pm(f_t-g))\|_p^p}\frac{\partial}{\partial t}\|\chi(\pm(f_t-g))\|_p^p, \text{ with } F_p(h) := \|\chi(h)\|_p^p$$

$$\begin{aligned}
\frac{\partial F_p}{\partial t}(\pm(f_t - g)) &= \frac{\partial}{\partial t}\int_{\mathbb{R}} |\chi(\pm(f_t-g))|^p \,\mathrm{d}\mu = \frac{\partial}{\partial t}\int_{\mathbb{R}} e^{\pm p(F_t(x)-G(x))}\,\mathrm{d}\mu(x)\\
&= \int_{\mathbb{R}} \frac{\partial}{\partial t} e^{\pm p(F_t(x)-G(x))}\,\mathrm{d}\mu(x)\\
&= \mp \int_{\mathbb{R}} p\left(\int_{x=-\infty}^{\cdot} f'(x-t)\,\mathrm{d}\mu(x)\right)\chi(\pm(f_t-g))^p\,\mathrm{d}\mu
\end{aligned}$$

and we finally get

$$\begin{aligned}
\frac{\partial \Gamma_p}{\partial t}(f_t - g) &= \frac{-1}{\|\chi(f_t-g)\|_p^p}\int_{\mathbb{R}}\left(\int_{x=-\infty}^{\cdot} f'(x-t)\,\mathrm{d}\mu(x)\right)\chi(f_t-g)^p\,\mathrm{d}\mu\\
&\quad + \frac{1}{\|\chi(-f_t+g)\|_p^p}\int_{\mathbb{R}}\left(\int_{x=-\infty}^{\cdot} f'(x-t)\,\mathrm{d}\mu(x)\right)\chi(-f_t+g)^p\,\mathrm{d}\mu\\
\frac{\partial \Gamma_p}{\partial t}(f_t - g) &= \left\langle \int_{x=-\infty}^{\cdot} f'(x-t)\,\mathrm{d}\mu(x), \left(\frac{\chi(-f_t+g)}{\|\chi(-f_t+g)\|_p}\right)^p - \left(\frac{\chi(f_t-g)}{\|\chi(f_t-g)\|_p}\right)^p\right\rangle
\end{aligned}$$

and this finishes the proof. □

Note that the theorem is stated for a continuous differentiable function f with no more constraints on the function g. A first remark is to notice that we could swap the constraints from f to g without changing the results, due to the commutativity of the convolution. As we stated earlier, f might be defined on a bounded domain only in which case the function f actually corresponds to the extension to the real line which might not be differentiable at the boundary of the domain. In this case, we can consider differentiating f in the sense of distribution which would add a Dirac δ distribution at the boundary for f'. This δ term added will have measure 0 for continuous measure or will be added to the total in case of a discrete measure. From a more functional analytic perspective, we could consider f to be a function of bounded variation and consider the first term in Eq. (2) to be the cumulative variation of f up to a certain point. We may equivalently rewrite:

$$\frac{\partial J}{\partial t}(t) \approx \left\langle \mathrm{CV}[f](\cdot - t)\,\mathrm{d}\mu(x), \left(\frac{\chi(-f_t+g)}{\|\chi(-f_t+g)\|_p}\right)^p - \left(\frac{\chi(f_t-g)}{\|\chi(f_t-g)\|_p}\right)^p\right\rangle$$

where $\mathrm{CV}[f](\cdot - t)$ corresponds, in the case of a differentiable function whose derivative is Riemann-integrable, to

$$\mathrm{CV}[f](x) = \int_{-\infty}^{x} f'(y)\mathrm{d}\mu(y)$$

While formula (2) seems complicated, we see that it simplifies in the case of real discrete signals.

Corollary 2 (Derivative for discrete uniform measure)**.** *Assume μ is a uniform discrete measure on a bounded domain. Then the derivative of the discrepancy autocorrelation function can be approximated by*

$$\frac{\partial J}{\partial t}(t) \approx \left\langle f_t, \left(\frac{\chi(-f_t+g)}{\|\chi(-f_t+g)\|_p}\right)^p - \left(\frac{\chi(f_t-g)}{\|\chi(f_t-g)\|_p}\right)^p\right\rangle,$$

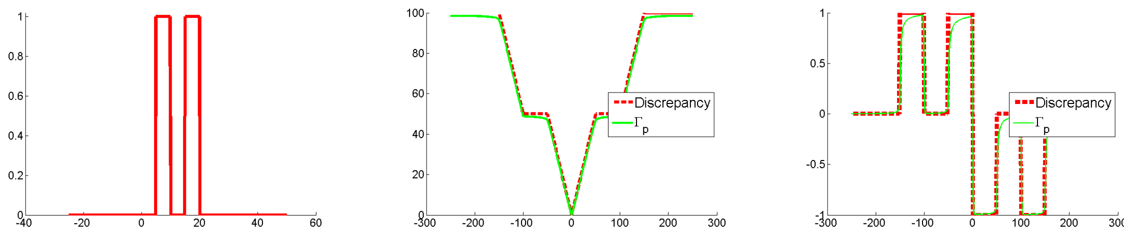
or equivalently:

$$\frac{\partial J}{\partial t}(t) \approx \sum_{i=1}^{N} f_t(i)\left(\left(\frac{\chi(-f_t+g)(i)}{\|\chi(-f_t+g)\|_p}\right)^p - \left(\frac{\chi(f_t-g)(i)}{\|\chi(f_t-g)\|_p}\right)^p\right),$$

where N denotes the size of the discrete finite signal (i.e.vector) we consider.

As stated earlier let us see on the toy example two steps function how the discrepancy norm can be problematic for local gradient based optimization algorithms. If we consider the autocorrelation function of the 2 steps example based on the discrepancy norm as well as on the Γ_p functions (see Fig. 4), it appears that even on really dummy easy example, the discrepancy norm does not appear to be well suited for optimization based procedure.

However, using an the above introduced approximations allows us to have an appropriate idea of what the discrepancy objective function looks like while keeping differentiable functions and avoiding plateaus which causes the derivative to vanish.



(a) Extended two steps function (b) Autocorrelation function: discrepancy and Γ_p (c) Approximate derivatives

Figure 4: This figure shows how a gradient descent based optimization algorithm might fail when trying to locally optimize the discrepancy correlation function. Indeed the correlation function (second figure) shows some plateau where the derivative (illustrated on the third figure) is 0. On the other hand, using an appropriate $p-$norm approximation allows to overcome this effect while keeping an really close objective function.

Now let us see how it behaves on some concrete examples. If we consider the horizontal wave function f and take it as a pattern we would like to align into a bigger pattern g, the discrepancy norm and its approximation(s) should ideally behave similarly when translating f along g. As toy examples, we consider two cases for g. The first one extends the f pattern only by padding 0s outside the domain of f. The second case mirrors the function f at its borders. These examples are illustrated in Figs. 5(a) to 5(c). The last Figure illustrates the case when the frequency of the wave is twice as big.

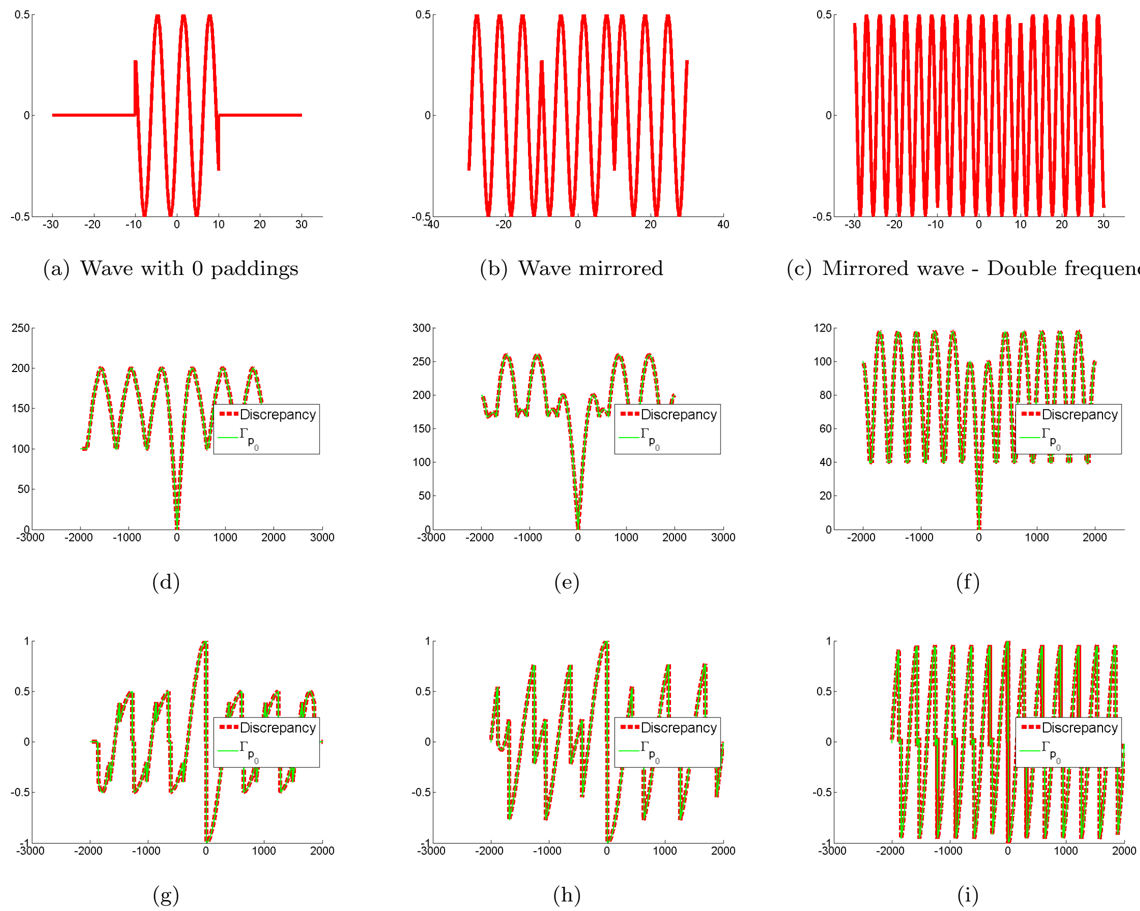
For a given function, we have computed the discrepancies between the local window of g and our pattern f. The same has been done for Γ_p where p has been chosen according to Theorem 2 using $\varepsilon = 0.1$ and a threshold of 0.2. The values are illustrated for displacement in the range of $[-N, N]$ for a uniform Dirac comb as measure in Figs. 5(d) to 5(f). It appears that the approximation overlap well with the original discrepancy autocorrelation function. Moreover, as we are aiming at minimizing this correlation function, we can remark that the different local minima and global minimum are located at the same places.

Finally, the last row of Fig. 5 shows two approximations of the discrepancy correlation. On is concerned with the analytical derivative of the Γ_p function over the shift, and the other one is computed as a finite difference of the discrepancy correlation: $\frac{\partial}{\partial t}\left(\Delta_{\|\cdot\|_D}[f](t)\right) \approx \Delta_{\|\cdot\|_D}[f](t+1) - \Delta_{\|\cdot\|_D}[f](t)$. See that not only is the approximated derivative fast to compute (only p-norms computations and a scalar product) we also get results pretty close to what numerical differentiation would give. Finally we can see that this approximation is suitable for any gradient based local optimization methods.

4 Approximation and optimization: the higher dimensional case

In this section we consider functions $f \in L(\Omega, \mathbb{R}; \mu)$ where:

- $\Omega \subseteq \mathbb{R}^d$ is open,
- d is the dimension of the ambient space,
- μ is a finite measure.



(a) Wave with 0 paddings (b) Wave mirrored (c) Mirrored wave - Double frequency

(d) (e) (f)

(g) (h) (i)

Figure 5: Sample sine wave with their discrepancy autocorrelation functions and their approximate derivatives. The first row shows the input signal extended with 0 padding (first column) and mirroring the wave (2^{nd} and 3^{rd} columns). 2^{nd} row shows the autocorrelation function based on the discrepancy norm and its approximation. Last row shows the approximated derivative together with the finite difference of the discrepancy autocorrelation.

We also consider the extended function $\widetilde{f}$ that matches f on Ω and is 0 elsewhere. Both f and $\widetilde{f}$ have the same discrepancy as introduced in the following sections.

4.1 Approximation of the discrepancy norm of a continuous multivariate signal

Now that the theory is clear for 1 dimensional functions, we would like to extend it to handle multivariate functions. According to [10], there are different ways of extending the discrepancy norm to higher dimensional functions but we consider only the following:

Definition 4 (Multivariate discrepancy norm)**.** *The discrepancy norm for multivariate functions can be defined by extending Eq. 3 to multidimensional integration. Let $\iota \in \{-1,1\}^d$ and to simplify the notations, let us define, for $s \in \mathbb{R}^d$, $[s[:=]-\infty, s_1[\times \cdots \times]-\infty, s_d[$. For a given ι, we denote by $\iota \cdot [s[$, the set composed of a product of interals: $\iota \cdot [s[:=]-\iota_1\infty, \iota_1 s_1[\times \cdots \times]-\iota_d\infty, \iota_d s_d[$, where the intervals might be flipped. Then we have:*

$$\|f\|_I^{(d)} := \max_{\iota \in \{-1,1\}^d} \left\{ \sup_{s \in \mathbb{R}^d} \int_{\iota \cdot [s[} f \, \mathrm{d}\mu - \inf_{s \in \mathbb{R}^d} \int_{\iota \cdot [s[} f \, \mathrm{d}\mu \right\}$$

And this equation can be numerically implemented by means of integral images [16, 17]. For simplicity, we use the following notation:

$$D^{(\iota)}(f) := \sup_{s\in\mathbb{R}^d} \int_{\iota\cdot[s[} f \,\mathrm{d}\mu - \inf_{s\in\mathbb{R}^d} \int_{\iota\cdot[s[} f \,\mathrm{d}\mu$$

It defines the discrepancy of a function given a direction of integration (by direction, we mean, from which corner of the space we start the integral image.)

Therefore we can equivalently write $\|f\|_I^{(d)} = \max_{1\leq i\leq 2^d}\{D^{(\iota^{(i)})}(f)\}$. We also define an approximation function through p-norms:

$$\Gamma_p^{(\iota)}(f) := \ln\left(M_p(\chi^{(\iota)}(f))\cdot M_p(\chi^{(\iota)}(-f))\right) = \ln\left(\frac{\|\chi^{(\iota)}(f)\|_p}{\mu(\mathbb{R}^d)^{1/p}}\cdot\frac{\|\chi^{(\iota)}(-f)\|_p}{\mu(\mathbb{R}^d)^{1/p}}\right)$$

with $\chi^{(\iota)}(f)(x) = e^{\int_{\iota\cdot[x[} f(s)\mathrm{d}\mu s}$. As in the previous sections, we have the following approximation theorem:

Theorem 5 (Approximation of the directional discrepancy norm multivariate functions)**.** $\Gamma_p^{(\iota)}$ *defines an approximation of the directional discrepancy norm in the sense that*

i) $\forall f\in L(\mathbb{R}^d;\mu),\ \Gamma_p^{(\iota)}(f) \underset{p\to\infty}{\longrightarrow} D^{(\iota)}(f)$

ii) $\Gamma_p(f)\leq D^{(\iota)}(f)$

iii) $\forall\varepsilon>0, \exists p^{(d,\iota)} : \forall p\in\mathbb{R}, p\geq p^{(d,\iota)} \Rightarrow |\Gamma_p^{(\iota)}(f) - D^{(\iota)}(f)|\leq\varepsilon$

Proof. The proof is the same as for Theorem 2, for continuous univariate functions.

However, for the last point, p^* has to be changed according to the dimension, with $\overline{\varepsilon} := 1 - e^{\varepsilon/2}$

$$p^{(d,\iota)} := \max\left\{\frac{\ln\frac{\mu(E_{\chi^{(\iota)}(f)}(t))}{\mu(\mathbb{R}^d)}}{\ln\frac{1-\overline{\varepsilon}}{1-t}}, \frac{\ln\frac{\mu(E_{\chi^{(\iota)}(-f)}(t))}{\mu(\mathbb{R}^d)}}{\ln\frac{1-\overline{\varepsilon}}{1-t}}\right\}$$

□

However, the problem is not solved yet. Indeed, we are able to smoothly approximate the directional discrepancies, but we still have to combine them by taking the max over all 2^d possible directions. Fortunately, this max over a finite set can be once again approximated with a reasonable error using $q-$norms:

$$\|f\|_D = \max_{\iota\in\{-1,1\}^d} D^{(\iota)}(f) \approx \|\overrightarrow{D}(f)\|_q$$

where $\overrightarrow{D} = \left[D^{(\iota^{(1)})},\cdots,D^{(\iota^{(2^d)})}\right]^T$ denotes the column vector composed with all directional discrepancies.

Finally we can combine both $p-q-$norms approximation:

$$\Gamma_{p,q}^{(d)}(f) := \left(\frac{\sum_{\iota\in\{-1,1\}^d}{\Gamma_p^{(\iota)}}^q(f)}{2^d}\right)^{1/q} = M_q(\overrightarrow{\Gamma_p})$$

where we have $\overrightarrow{\Gamma_p} = \left[\Gamma_p^{(\iota^{(1)})},\cdots,\Gamma_p^{(\iota^{(2^d)})}\right]^T$

Corollary 3 (Convergence of $\Gamma_{p,q}$)**.** *Given a function $f\in L(\mathbb{R}^d,\mathbb{R},\mu)$, the following holds:*

$$\forall\varepsilon>0, \exists\mathbf{p}^{(d)} := (p^{(d)},q^{(d)}) : \forall\mathbf{p}\geq\mathbf{p}^{(d)}, \left|\Gamma_{p,q}^{(d)}(f) - \|f\|_I^{(d)}\right|\leq\varepsilon$$

Note that $\mathbf{p} = (p,q)\geq\mathbf{p}^{(d)}$ reads $p\geq p^{(d)}$ and $q\geq q^{(d)}$.

Proof. Let $0 < \lambda < 1$ and $\varepsilon > 0$.

We need to estimate the quantity $\delta_{p,q} := \left| \|f\|_I^{(d)} - \Gamma_{p,q}^{(d)}(f) \right|$.

$$\delta_{p,q} = \|f\|_I^{(d)} - M_q(\overrightarrow{D}(f)) + M_q(\overrightarrow{D}(f)) - \Gamma_{p,q}^{(d)}(f) =: d_1 + d_2$$

Now let us start with the first component

$$d_1 = \|f\|_I^{(d)} - M_q(\overrightarrow{D}(f)) = \|\overrightarrow{D}(f)\|_\infty - M_q(\overrightarrow{D}(f))$$

this last difference actually lies in a finite-dimensional vector space (2^d dimensions) and we have the following norm comparisons:

$$\|\overrightarrow{D}(f)\|_\infty \leq \|\overrightarrow{D}(f)\|_q \leq 2^{d/q}\|\overrightarrow{D}(f)\|_\infty$$

which means that for q large enough ($q \geq q_0$), we have $\left| \|\overrightarrow{D}(f)\|_q - \|\overrightarrow{D}(f)\|_\infty \right| < (1-\lambda)\varepsilon$. Indeed

$$\|\overrightarrow{D}(f)\|_\infty \leq \|\overrightarrow{D}(f)\|_q \leq 2^{d/q}\|\overrightarrow{D}(f)\|_\infty$$

is equivalent to

$$\frac{\|\overrightarrow{D}(f)\|_\infty}{2^{d/q}} \leq M_q(\overrightarrow{D}(f)) \leq \|\overrightarrow{D}(f)\|_\infty$$

which can be reformulated as

$$\frac{\|\overrightarrow{D}(f)\|_\infty}{2^{d/q}} - \|\overrightarrow{D}(f)\|_\infty \leq M_q(\overrightarrow{D}(f)) - \|\overrightarrow{D}(f)\|_\infty \leq 0$$

whence for $-(1-\lambda)\varepsilon \leq \|\overrightarrow{D}(f)\|_\infty \left(2^{-d/q} - 1\right)$ we get the results.

$$1 - \frac{(1-\lambda)\varepsilon}{\|\overrightarrow{D}(f)\|_\infty} \leq 2^{-d/q}$$

$$\frac{d\ln(2)}{\ln\left(\frac{\|\overrightarrow{D}(f)\|_\infty}{\|\overrightarrow{D}(f)\|_\infty - (1-\lambda)\varepsilon}\right)} \leq q$$

And we set $q_0 := d\ln 2/\ln\left(\frac{\|f\|_I^{(d)}}{\|f\|_I^{(d)} - (1-\lambda)\varepsilon}\right)$

Now let us compute an estimate for the second part fixing $q = q_0$, and try to get $d_2 \leq \lambda\varepsilon$ to prove the corollary:

$$d_2 = M_q(\overrightarrow{D}(f)) - \Gamma_{p,q}^{(d)}(f) = M_q(\overrightarrow{D}(f)) - M_q(\overrightarrow{\Gamma_p}(f)) = \frac{\|\overrightarrow{\Gamma_p}(f)\|_q - \|\overrightarrow{D}(f)\|_q}{2^{d/q}} \leq \lambda\varepsilon$$

Thus applying the estimates of the triangular inequalities for the q-norm, it suffices to find p such that

$$\forall \iota \in \{-1,1\}^d, D^{(\iota)}(f) - \Gamma_p^{(\iota)}(f) \leq \lambda\varepsilon$$

which implies $\displaystyle\sum_{\iota\in\{-1,1\}^d} \left| D^{(\iota)}(f) - \Gamma_p^{(\iota)}(f) \right|^p \leq 2^d\lambda^q\varepsilon^q$ and consequently $\|\overrightarrow{D}(f) - \overrightarrow{\Gamma_p}(f)\|_q \leq 2^{d/q}\lambda\varepsilon$

Now we get that for all ι, p has to be greater to $p^{(\iota)}$ according to Theorem 5. In other terms, by defining $p_0 := \max_\iota p^{(\iota)}$ it ensures that $d_2 \leq \lambda\varepsilon$ and therefore we have $\delta_{p,q} \leq (1-\lambda)\varepsilon + \lambda\varepsilon$ which finishes the proof. □

Remark that the most important part of the approximation, and the smallest one we need to get, is the one regarding p. The one regarding q is numerically more stable. Moreover, q is just a classical 2^d dimensional vector norm which will be used for d being practically 2 (for images) of 3 (for volumes) and does not create troubles for applications. Therefore we might suggest to get λ as close to 1 as possible.

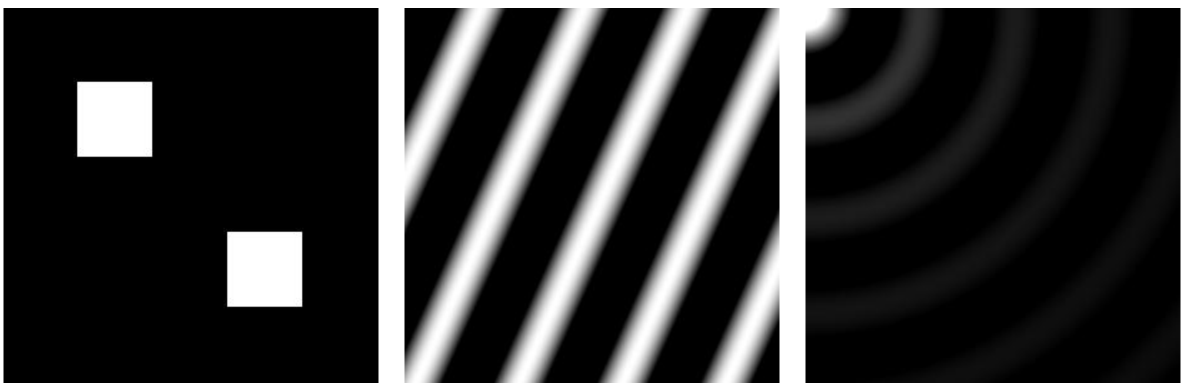
Another remark comes from the computation of q_0 where the value of the discrepancy norm of f is actually needed. In this case one can question itself whether all these calculations make sense if we anyway have to compute the discrepancy norm to get an estimator of it. However in the practical example concerned with autocorrelation, we estimate this value only once for the whole registration procedure.

Examples of lower bounds Here we introduce three toy examples of functions defined on $\mathbb{R}^2$, which we use as test samples.

$$f_1 : \begin{cases} [1,50]\times[1,50] & \to & \mathbb{R} \\ \mathbf{x}=(x,y) & \mapsto & 1_{[10,20[} + 1_{[30,40[} \end{cases}$$
$$f_2 : \begin{cases} [1,50]\times[1,50] & \to & \mathbb{R} \\ \mathbf{x}=(x,y) & \mapsto & \sin(\mathbf{u}^T\mathbf{x}) \end{cases}$$
$$f_3 : \begin{cases} [1,50]\times[1,50] & \to & \mathbb{R} \\ \mathbf{x}=(x,y) & \mapsto & \frac{1}{\|\mathbf{x}\|_2}\sin(\|\mathbf{x}\|_2) \end{cases}$$

In the definition of f_2 the vector $\mathbf{u}$ is characterized by two parameters: an angle α and a period T, so that we can also write $f_2(x,y) = \sin\left(\frac{\cos(\alpha)x+\sin(\alpha)y}{T}\right)$

These examples are depicted on Fig. 6. It must be noted that the second and first images (or functions) are oscillating around 0 while the first one is nonnegative. This has a particular effect as it can be seen in the results of Table. 2.



(a) 2 blocs function (b) Sine waves (c) Radial waves

Figure 6: Our dataset of 2 dimensional toy functions used in our paper.

Table. 2 gathers some results of the convergence and quality of the approximation for the different test functions with different parameters. It is divided into three groups of columns: the first one describes the different parameters used for the experiments (function, the maximal error expected ε, the percentage of ε needed for the thresholding process in the computation of p_0 $\alpha \in [0,1]$ and the repartition of the error between p-mean and q-mean approximations $\lambda \in [0,1]$), the results obtained for with our approximations and estimations (p_0 and q_0 computed according to Corollary 3, the value of the approximation Γ_{p_0,q_0} and the value of the discrepancy norm) and the optimal exponents $\overline{p_0}$, $\overline{q_0}$ achievable with the given the current set of parameters. They are computed as follows:

$$\overline{q_0} = \min\left\{q \in \mathbb{N} : \|f\|_I^{(2)} - M_q(\overrightarrow{D}(f)) < (1-\lambda)\varepsilon\right\}$$
$$\overline{p_0} = \min\left\{p \in \mathbb{N} : M_{\overline{q_0}}(\overrightarrow{D}(f)) - M_{\overline{q_0}}(\overrightarrow{\Gamma}_p(f)) < \lambda\varepsilon\right\}$$

It is important to us that the estimated p_0 and q_0 stay close to the optimal values $\overline{p_0}$ and $\overline{q_0}$ so that the quality of the approximation if preserved, without getting into numerical troubles.

Concerning the first example of the two blocs functions, we see that our estimates are far above the optimal ones. This is due to the fact that the function is non negative. In this case, $\forall q \in \mathbb{R}, M_q(\overrightarrow{D}(f)) = \|\overrightarrow{D}(f)\|_\infty = \min(\overrightarrow{D}(f)) = \|f\|_I^{(2)}$ so that no error is done on any non negative functions during the q-means approximation.

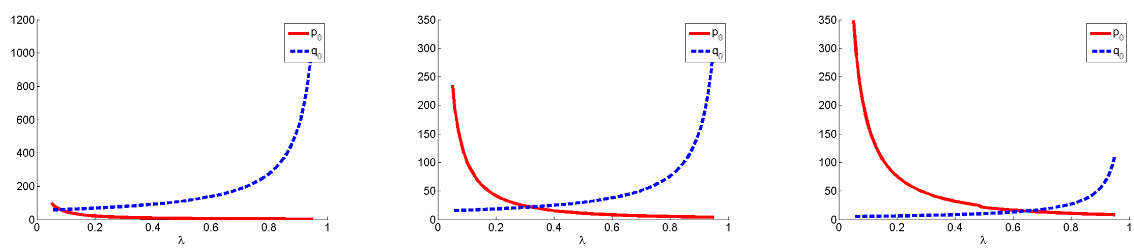
The second function is an example of a (at least locally, due to the finiteness of the image) periodic function. In this case we notice that our estimation are close to the optimal ones (sometimes even smaller due to the discretization of the computation of $\overline{q_0}$ and $\overline{p_0}$). Why this happens to be for such periodic signals has to be studied into more details and goes beyond the scope of this research.

In the last example of spherical waves, we can see that estimated powers are within two to three times bigger than the optimal ones yielding an error about half the maximal error expected.

Function	ε	α	λ	p_0	q_0	Γ_{p_0,q_0}	$\|\cdot\|_I^{(2)}$	$\overline{p_0}$	$\overline{q_0}$
Blocs	1	0.75	0.75	29.88	1108.3	199.87	200	6	1
	1	0.75	0.5	46.92	553.82	199.92	200	6	1
	1	0.75	0.25	98.30	368.98	199.96	200	6	1
	1	0.5	0.5	24.24	553.82	199.84	200	6	1
	10	0.5	0.5	1.71	54.76	197.73	200	1	1
Sines	1	0.25	0.5	40.52	154.09	54.97	55.83	29	155
	1	0.3	0.5	39.32	154.09	54.96	55.83	29	155
	1	0.5	0.5	53.70	154.09	55.05	55.83	29	155
	1	0.75	0.5	90.59	154.09	55.16	55.83	29	155
	2.5	0.3	0.5	13.56	61.22	53.63	55.83	11	60
Radial	0.25	0.5	0.5	246.52	179.30	16.10	16.23	125	90
	0.5	0.5	0.5	121.39	89.30	15.98	16.23	62	42
	1	0.5	0.5	58.92	44.30	15.73	16.23	31	17
	1	0.5	0.4	74.52	36.80	15.75	16.23	38	13

Table 2: Examples of approximation and lower bound estimations on some toy functions.

What come out of this Table is that the choice of an adequate λ is vital to avoid numerical problems. Indeed, if not well tuned, it leads to high p or q powers and yields numerical overflows. Fig. 7 shows the impact of this λ on the estimated p_0 and q_0.



(a) 2 blocs function (b) Sine waves (c) Radial waves

Figure 7: Evolution of the values of p_0 and q_0 depending on the choice of λ.

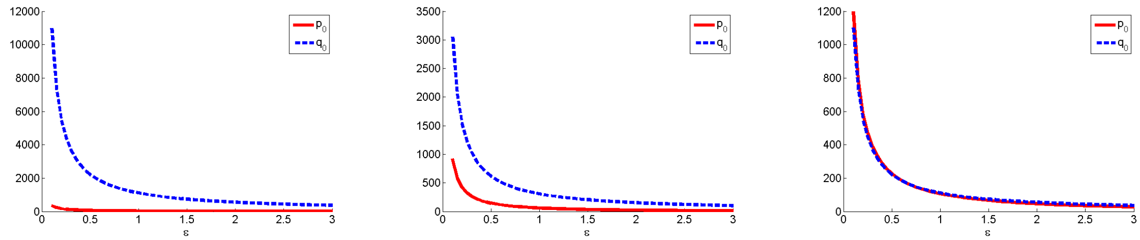
It seems that, apart from the blocs samples, a λ within the range $[0.3, 0.7]$ gives a good compromise between a high q or a high p.

4.2 Differentiation of the discrepancy correlation: the multivariate case

As in the one dimensional case, we introduce the discrepancy correlation function which we wish to minimize $J^{(d)}(\mathbf{t}) := \|f_{\mathbf{t}} - g\|_I^{(d)}$, where f and g denotes two patterns we want to align.

Theorem 6 (Gradient computation of the discrepancy correlation for multivariate differentiable functions)**.** *Let f be continuously differentiable and measurable function and g be measurable. Let $J^{(d)}(\mathbf{t}) := \|f_{\mathbf{t}} - g\|_I^{(d)}$ denote the discrepancy correlation function. Then the derivatives of $J^{(d)}(\boldsymbol{t})$ can be approximated in the following way:*

$$\frac{\partial J}{\partial \boldsymbol{t}_i}(t) \approx \frac{1}{2^{d/q}} \frac{\|\overrightarrow{\Gamma}_p(f_{\boldsymbol{t}} - g)\|_q}{\|\overrightarrow{\Gamma}_p(f_{\boldsymbol{t}} - g)\|_q^q} \sum_{\iota \in \{-1,1\}^d} \Gamma_p^{(\iota)}(f_{\boldsymbol{t}} - g)^{q-1} \left\langle \theta_i^{(\iota)}(\cdot - \boldsymbol{t}), \delta_{\chi^{(\iota)}} \right\rangle \tag{3}$$



(a) 2 blocs function (b) Sine waves (c) Radial waves

Figure 8: Evolution of the values of p_0 and q_0 depending on the choice of ε.

where we use the following notations:

$$\theta_i^{(\iota)}(x-\mathbf{t}) = \int_{\iota\cdot[x[} \frac{\partial f}{\partial \boldsymbol{x}_i}(s-t)\mathrm{d}\mu(s) \tag{4}$$

$$\delta_{\chi^{(\iota)}} = \left(\frac{\chi^{(\iota)}(g-f_{\mathbf{t}})}{\|\chi^{(\iota)}(g-f_{\mathbf{t}})\|_p}\right)^p - \left(\frac{\chi^{(\iota)}(f_{\mathbf{t}}-g)}{\|\chi^{(\iota)}(f_{\mathbf{t}}-g)\|_p}\right)^p$$

Note that we consider here a function f defined on $\mathbb{R}^d$. In the case that f is only defined on a domain $\Omega \subsetneq \mathbb{R}^d$ we would expect the function to be smooth at the boundary such that its extension on the whole space is continuously differentiable. We can extend these results by considering function in a Sobolev space $W^{1,1}(\Omega)$ instead.

Proof. Let $\Delta_{p,q}[f,g](\mathbf{t}) := \Gamma_{p,q}^{(d)}(f_{\mathbf{t}}-g) = \frac{\left(\sum_\iota \Gamma_p^{(\iota)}(f_{\mathbf{t}}-g)^q\right)^{1/q}}{2^d}$. We approximate $J^{(d)}(\mathbf{t}) \approx \Delta_{p,q}[f,g](\mathbf{t})$.

Let $i \in \{1,\cdots,d\}$ we have

$$\begin{aligned}
\frac{\partial \Delta_{p,q}[f,g]}{\partial \mathbf{t}_i}(\mathbf{t}) &= \frac{1}{2^{d/q}}\frac{1}{q}\left(\sum_\iota \Gamma_p^{(\iota)}(f_{\mathbf{t}}-g)^q\right)^{1/q-1} \frac{\partial \sum_\iota \Gamma_p^{(\iota)}(f_{\mathbf{t}}-g)^q}{\partial \mathbf{t}_i} \\
&= \frac{1}{2^{d/q}}\frac{1}{q}\frac{\|\overrightarrow{\Gamma}_p(f_{\mathbf{t}}-g)\|_q}{\|\overrightarrow{\Gamma}_p(f_{\mathbf{t}}-g)\|_q^q}\frac{\partial \sum_\iota \Gamma_p^{(\iota)}(f_{\mathbf{t}}-g)^q}{\partial \mathbf{t}_i} \\
\frac{\partial \sum_\iota \Gamma_p^{(\iota)}(f_{\mathbf{t}}-g)^q}{\partial \mathbf{t}_i} &= q\sum_\iota \Gamma_p^{(\iota)}(f_t-g)^{q-1}\frac{\partial \Gamma_p^{(\iota)}(f_t-g)}{\partial \mathbf{t}_i} \\
\frac{\partial \Gamma_p^{(\iota)}(f_{\mathbf{t}}-g)}{\partial \mathbf{t}_i} &= \left\langle \theta_i^{(\iota)}(\cdot-\mathbf{t}), \delta_{\chi^{(\iota)}}\right\rangle
\end{aligned}$$

So that combining every components together we get

$$\frac{\partial \Delta_{p,q}[f,g]}{\partial \mathbf{t}_i}(\mathbf{t}) = \frac{1}{2^{d/q}}\frac{\|\overrightarrow{\Gamma}_p(f_{\mathbf{t}}-g)\|_q}{\|\overrightarrow{\Gamma}_p(f_{\mathbf{t}}-g)\|_q^q}\sum_{\iota\in\{-1,1\}^d}\Gamma_p^{(\iota)}(f_{\mathbf{t}}-g)^{q-1}\left\langle \theta_i^{(\iota)}(\cdot-\mathbf{t}), \delta_{\chi^{(\iota)}}\right\rangle$$

□

4.3 Discussions

Implementation. While Eq. (3) seems quite complicated and not really suitable for practical use. We first look at and decompose all the different components of the formula. The first one $\frac{1}{2^{d/q}}\frac{\|\overrightarrow{\Gamma}_p(f_{\mathbf{t}}-g)\|_q}{\|\overrightarrow{\Gamma}_p(f_{\mathbf{t}}-g)\|_q^q}$ is just a multiplicative factor and can be simplified as $2^{-d}\Gamma_{p,q}(f_{\mathbf{t}}-g)^{1-q}$ which depends neither on i nor ι.

The sum on the right can be understood as a scalar product in $\mathbb{R}^{2^d}$. As ι travels all over $\{-1,1\}^d$, each $\Gamma_p(\iota)$ takes value in $\mathbb{R}$ (and actually only positive values as already seen). So does the scalar product (in $\mathbb{R}^d$ this time) $\left\langle \theta_i^{(\iota)}(\cdot - \mathbf{t}), \delta_{\chi^{(\iota)}} \right\rangle$; so that alltogether the sum can be interpreted as $\left(\overrightarrow{\Gamma_p(f_{\mathbf{t}} - g)}^{q-1}, \overrightarrow{\Theta_i}(\mathbf{t}) \right)$ where $\overrightarrow{\Theta_i}(t) = \left[\int_{x\in\mathbb{R}^d} \theta_i^{(\iota)}(x-\mathbf{t})\delta_{\chi^{(\iota)}}(x)\mu(x) \right]_{\iota\in\{-1,1\}^d}$. The power of vector $\overrightarrow{\Gamma_p}$ is to be understood componentwise.

In many practical cases, the μ measure will be slowly varying (if not completely uniform on a compact support) and discrete. In this configuration, and in particular in the case of image, $d=2$, we have (assuming $\iota = (1,1)$ i.e., the summation start from the $-\infty,-\infty$ corner; the other cases are computed similarly)

$$\begin{aligned}
\theta_i^{(\iota)}(x-\mathbf{t}) &= \int_{\iota\cdot[x[} \frac{\partial f}{\partial \mathbf{t}_i}(s-t)\mathrm{d}\mu(s) \\
&\approx \sum_{l=-\infty}^{x_1} \sum_{m=-\infty}^{x_2} \frac{\partial f}{\partial \mathbf{t}_i}(l-t_1, m-t_2)\mu(l,m)
\end{aligned}$$

The partial derivative can be approximated by finite differences; in the continuous case, the fundemental theorem of calculus applies.

$$\begin{aligned}
\theta_i^{(\iota)}(x-\mathbf{t}) &= \sum_{l=-\infty}^{x_1} \sum_{m=-\infty}^{x_2} \left(f(l-t_1+1, m-t_2) - f(l-t_1, m-t_2) \right) \mu(l,m) \\
&\approx \sum_{m=-\infty}^{x_2} f(x_1+1-t_1, m-t_2)\mu(x_1+1, m)
\end{aligned}$$

and the last formula tells that we only need to do some kind of integral images in $n-1$ direction (apart from the one in which we wish to differentiate). This is easy and efficient to compute when working on images.

Non-differentiability of f. Theorem 6 requires f to be continuously differentiable. The only reason to have this is to have $\theta_i^{(\iota)}$ in Eq. (4) to be well defined. We could also consider f to be only μ-almost everywhere differentiable. In this case we have that $\frac{\partial f}{\partial \mathbf{x}_i}(s-t)$ exists almost everywhere and hence Eq. (4) is well defined. Evidence suggest that similar results exist for function of bounded variations defined on Ω but the proof of this statement being more technical is left aside for this paper.

5 Conclusion

We have reintroduced here the use of the discrepancy norm as a tool to measure correlations and (dis-) similarities between finite-length (or finite energy) signals. We proved that this norm can be approximated as close as we want by L^p norms and that this norm can be used in gradient based optimization procedures for, for instance, alignment purposes. Moreover, calculations may be accelerated using a convolution formula that take advantage of this new approximation.

As these results hold also in high-dimensional signals, we believe that our results could be applied to image registration, for instance using a log-polar transform [21, 22], or to any phase-based alignment.

Some care should however be taken when computing this approximation as values may get very high very soon with the degree of approximation increasing.

Acknowledgements

This work was initiated while JLB was funded by the Austrian Science Fund (FWF, grant no. P21496 N23, while he was with Department of Knowledge-based Mathematical System, Johannes Kepler University, Linz, Austria). He thanks now the National Science Foundation (grant DMS-1120622) for financial support. The authors are thankful to the anonymous reviewers for their precious comments.

References

[1] H. Weyl, Über die Gleichverteilung von Zahlen mod. Eins, Math. Ann. 77 (1916) 313–352.

[2] S. Haber, Numerical evaluation of multiple integrals, SIAM Review 12 (4) (1970) 481–526.

[3] R. Coveyou, R. D. MacPherson, Fourier analysis of uniform random number generators, Journal of the ACM 14 (1967) 100–119.

[4] J. Koksma, Ein mengentheoretischer satz über die gleichverteilung modulo eins, Compositio Mathematica 2 (1935) 250–258.

[5] H. Niederreiter, Random number generation and quasi-Monte Carlo methods, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.

[6] B. Chazelle, The discrepancy method in computational geometry, in: In Handbook of Discrete and Computational Geometry, CRC Press, 2004, pp. 983–996.

[7] J. Matoušek, Geometric discrepancy: An illustrated guide, Vol. 18, Springer, 1999.

[8] B. Chazelle, The discrepancy methods: Randomness and complexity, Cambridge University Press, 2002.

[9] H. Neunzert, B. Wetton, Pattern recognition using measure space metrics, Tech. Rep. 28, Universitt Kaiserslautern, Fachbereich Mathematik (November 1987).

[10] B. Moser, A similarity measure for image and volumetric data based on hermann weyl's discrepancy, Pattern Analysis and Machine Intelligence, IEEE Transactions on 33 (11) (2011) 2321 –2329. doi:10.1109/TPAMI.2009.50.

[11] J.-L. Bouchot, J. Himmelbauer, B. Moser, On autocorrelation based on Hermann Weyl's discrepancy norm for time series analysis, in: A. Prieto (Ed.), Neural Networks (IJCNN), The 2010 International Joint Conference on, IEEE, 2010, pp. 3139 – 3145. doi:10.1109/IJCNN.2010.5596843.

[12] B. Moser, G. Stübl, J.-L. Bouchot, On a non-monotonicity effect of similarity measures, Similarity-Based Pattern Recognition (2011) 46–60.

[13] J.-L. Bouchot, G. Stübl, B. Moser, A template matching approach based on the discrepancy norm for defect detection on regularly textured surfaces, in: 10th International Conference on Quality Control by Artificial Vision, International Society for Optics and Photonics, 2011, pp. 80000K–80000K.

[14] G. Stübl, J.-L. Bouchot, P. Haslinger, B. Moser, Discrepancy norm as fitness function for defect detection on regularly textured surfaces, in: Pattern Recognition, Springer, 2012, pp. 428–437.

[15] P. Bauer, U. Bodenhofer, E. P. Klement, A fuzzy algorithm for pixel classification based on the discrepancy norm, in: Proc. of the 5th IEEE Int. Conf on Fuzzy Syst., Vol. III, New Orleans, LA, 1996, pp. 2007–2012.

[16] F. Crow, Summed-area tables for texture mapping, ACM SIGGRAPH Computer Graphics 18 (3) (1984) 207–212.

[17] P. Viola, M. Jones, Robust real-time object detection, International Journal of Computer Vision 57 (2) (2001) 137–154.

[18] A. Kolmogorov, G. Castelnuovo, Sur la notion de la moyenne, G. Bardi, tip. della R. Accad. dei Lincei, 1930.

[19] G. Hardy, J. Littlewood, G. Pólya, Inequalities, Cambridge Univ Pr, 1952.

[20] L. Loomis, An introduction to abstract harmonic analysis, Van Nostrand Reinhold, 1953.

[21] G. Wolberg, S. Zokai, Robust image registration using log-polar transform, in: Image Processing, 2000. Proceedings. 2000 International Conference on, 2000, pp. 493 –496 vol.1. doi:10.1109/ICIP.2000.901003.

[22] S. Zokai, G. Wolberg, Image registration using log-polar mappings for recovery of large-scale similarity and projective transformations, Image Processing, IEEE Transactions on 14 (10) (2005) 1422 –1434. doi:10.1109/TIP.2005.854501.