Introducing a new family of metrics for comparing two finite sets of $\mathscr{L}^2$ functions

Octavian Neamtu

December 22, 2011

1 Introduction

1.1 Statement of purpose

Our aim is to introduce a family of metrics that can be used to find how similar two finite ordered sets of $\mathscr{L}^2$ functions on some compact set $X \subset \mathbb{R}$. Each of the two sets of functions has the same number of elements and the comparison is done one-by-one. So if the sets were $A = \{f_1, f_2, \ldots f_n\}$ and $B = \{g_1, g_2, \ldots g_n\}$, we would like to compare f_1 against g_1, f_2 against g_2, etc.

The reason why we choose our functions to be $\mathscr{L}^2(X)$ is because $\mathscr{L}^2(X)$ naturally forms a Hilbert space with the dot product

$$\langle f, g\rangle = \int_X f g d\mu$$

and the dot product provides a primal way of comparing two functions.

One of the applications that motivate this study is the problem of finding similarities between some set number stock price functions at critical points in time such as stock market crashes. We make the simplifying assumption that the stock prices can be accurately modeled with geometric Brownian motion functions. Since we are operating on a compact space in $\mathbb{R}$, the continuity of these functions ensure the fact that they are $\mathscr{L}^2$ functions. The other simplifying assumption we make is that one can compute the integral of a geometric Brownian motion; this is impossible to do in a direct manner, but can be overcome by using methods from stochastic calculus.

Definition 1. *Given n functions then consider a function matrix as a matrix where each row is replaced by a function.*

$$A = \begin{pmatrix} -f_1(x)- \\ -f_2(x)- \\ \cdots \\ -f_n(x)- \end{pmatrix}$$

Note. *A is also an operator on the $\mathscr{L}^2$ space to $\mathbb{R}^n$ and for any $x \in \mathscr{L}^2$ we have that*

$$Ax = \begin{pmatrix} \langle f_1, x\rangle \\ \langle f_2, x\rangle \\ \cdots \\ \langle f_n, x\rangle \end{pmatrix}$$

where $\langle \cdot, \cdot \rangle$ is the $\mathscr{L}^2$ inner product. Label this operator space as O_f.

Note. *A is a function from $\mathbb{R}^n$ to $\mathbb{R}^n$ in the sense that*

$$A(y) = \begin{pmatrix} f_1(y) \\ f_2(y) \\ \dots \\ f_n(y) \end{pmatrix}$$

Note. *If we took the n functions to be discrete functions defined on set of k elements then A is nothing more than a $n \times k$ real-valued matrix.*

1.2 Acknowledged previous work

Yu, Houle and Mio[1] extended the usual distance metric used in Procrustean shape analysis to a family of metrics that would yield a better result in the comparison of shapes when the appropriate metric was chosen. The problem they dealt with was comparing a finite set of landmark points in two images using the extended metric and check whether the shapes each of the point sets determined were similar or not.

In a similar sense, this paper proposes to compare two entities, but instead of shapes determined by a finite number of 2D points, we deal with a more abstract sense of shape that is determined by an n-space of $\mathscr{L}^2$ functions. We try to approach the problem in the same way Yu, Houle and Mio[1], but the problems that arise are of a very different nature.

2 The new norm on the block

2.1 The basis for the new norm

If our functions were discrete as described above, then to compare two such matrices one's first intuition is to use the distance $||A - B||_F$, where $|| \cdot ||_F$ is the Frobenius norm. This distance is also the basis of the Procrustean shape analysis methodology used by Yu, Houle and Mio [1]. But since we are dealing with a Hilbert space rather than a discrete space, we will use the natural generalization of the Frobenius norm for operators on a Hilbert space called the Hilbert-Schmidt norm. The problem that arises is that even though all matrices have a finite Frobenius norm because of their inherent finiteness, an operator does not necessarily have a finite Hilbert-Schmidt norm.

Definition 2. *An operator is Hilbert-Schmidt $\iff$ the Hilbert-Schmidt norm of the operator is finite.*

Definition 3. *The Hilbert-Schmidt norm of an operator on a Hilbert space is defined as*

$$||A||_{HS} = \sum_{i,j} |\langle e_i, Ae_j \rangle|^2 = \sum_j ||Ae_j||^2$$

where e_i are an orthonormal basis for the underlying Hilbert space. The Hilbert-Schmidt distance then is $||A - B||_{HS}$.

Lemma 1. *The Hilbert-Schmidt norm dominates the operator 2-norm. That is, for any Hilbert space E and $x \in E$*

$$||Tx||_2 \leq ||x||_2||T||_{HS}$$

Proof. Take any Hilbert space E and e_i an orthonormal base such that for any $x = \sum_j c_j e_j \in E$ we have that

$$\begin{aligned}
||Tx||_2^2 &= ||T(\sum_j c_j e_j)||_2^2 \\
&= ||\sum_j c_j T e_j||_2^2 \\
&\leq \sum_j ||c_j T e_j||_2^2 \\
&= \sum_j |c_j|^2 ||Te_j||_2^2 \\
&\leq (\sum_j |c_j|^2)(\sum_j ||Te_j||_2^2) \\
&= ||x||_2^2 ||T||_{HS}^2
\end{aligned}$$

□

The Hilbert-Schmidt norm is then in some sense more comprehensive than the usual operator 2-norm because it dominates the latter. This makes the Hilbert-Schmidt distance be more sensitive to differences between the two operators that the 2-norm distance.

To check that our operator is Hilbert-Schmidt, we first check that is bounded with the extended 2-norm from $\mathbb{R}^n$ onto our operator space and then check that it is Hilbert-Schmidt.

Lemma 2. *Any operator in the operator space O_f is bounded in the operator 2-norm.*

Proof.

$$||A||_2 = \sup\{||Px||_2 \mid ||x||_2 = 1\}$$

or in other words

$$||A||_2 = \sup \sqrt{\sum_{i=1}^{n} \langle f_i, x \rangle^2}$$

For our operator to be bounded we must have that

$$||A||_2 < \infty$$

But from the *Cauchy-Schwartz* inequality we know that

$$\langle f_i, x \rangle^2 \leq ||f_i||^2 ||x||^2 = ||f_i||^2$$

Then

$$\sum_{i=1}^{n}\langle f_i, x\rangle^2 \leq \sum_{i=1}^{n}||f_i||^2$$

and since f_i is an $\mathscr{L}^2$ function and therefore has a finite norm, we get that the sum of the norms is also finite. Therefore

$$||A||_2 = \sup\sqrt{\sum_{i=1}^{n}\langle f_i, x\rangle^2} \leq \sum_{i=1}^{n}||f_i||^2 < \infty$$

□

Lemma 3. *Any operator in the operator space O_f is Hilbert-Schmidt.*

Proof. From the definition of the Hilbert-Schmidt norm on an operator space, we get that

$$\begin{aligned}
||A||_{HS}^2 &= \sum_j ||Ae_j||^2 \\
&= \sum_j \sum_{i=1}^{n}\langle f_i, e_j\rangle^2 \\
&= \sum_{i=1}^{n}\sum_j \langle f_i, e_j\rangle^2 \\
&= \sum_{i=1}^{n}||f_i||^2
\end{aligned}$$

which we know is finite from the proof of *Lemma 2*. □

Note. *From the proofs of Lemma 2 and Lemma 3 we can verify that indeed*

$$||A||_2 \sup\sqrt{\sum_{i=1}^{n}\langle f_i, x\rangle^2} \leq \sum_{i=1}^{n}||f_i||^2 = ||A||_{HS}$$

Note. *The Hilbert-Schmidt operators form a Hilbert space with the inner product*

$$\langle A, B\rangle_{HS} = \sum_i \langle Ae_i, Be_i\rangle$$

For operators in O_f this inner product is

$$\begin{aligned}\langle A, B\rangle_{HS} &= \sum_i \langle Ae_i, Be_i\rangle \\ &= \sum_i \langle \sum_{j=1}^{n} \langle f_j, e_i\rangle \langle g_j, e_i\rangle\rangle \\ &= \sum_{j=1}^{n} \langle \sum_i \langle f_j, e_i\rangle \langle g_j, e_i\rangle\rangle \\ &= \sum_{j=1}^{n} \langle f_i, g_i\rangle\end{aligned}$$

It is also interesting to note that the dot product we obtain is in fact the sum of the pair-wise dot product of the functions we wanted to compare.

2.2 Extending the Hilbert-Schmidt norm to a new family of norms

The norm we propose adds a new layer of flexibility by adding a multiplication operator to the mix. This multiplication operator can highlight critical parts of the function $A(x)$ by multiplying it with a high value on those intervals or dampen other parts of it by multiplying the function with a low value on those intervals. This is similar to the symmetric matrix that Yu, Houle and Mio [1] use to highlight certain landmark points in their metric, and in some sense is the operator equivalent of a diagonal matrix.

Definition 4. *For M_ϕ a multiplication operator from $\mathscr{L}^2$ to $\mathscr{L}^2$ with some function ϕ, let*

$$A_\phi = AM_\phi$$

Lemma 4. *A_ϕ is an operator in O_f.*

Proof.

$$(AM_\phi)x = A(M_\phi)x = A\phi x = \begin{pmatrix} \cdots \\ \langle f_i, \phi x\rangle \\ \cdots \end{pmatrix} = \begin{pmatrix} \cdots \\ \langle f_i \phi, x\rangle \\ \cdots \end{pmatrix} = A_\phi x$$

Since M_ϕ is an operator from $\mathscr{L}^2$ to $\mathscr{L}^2$ then $f_i\phi$ are all $\mathscr{L}^2$ functions and thus A_ϕ is in O_f. □

Lemma 5. *If ϕ is bounded by some finite B, then M_ϕ is an operator from $\mathscr{L}^2$ to $\mathscr{L}^2$.*

Proof. In order to prove that M_ϕ is an operator from $\mathscr{L}^2$ to $\mathscr{L}^2$ then we need to prove that $M_\phi x$ is still an $\mathscr{L}^2$ function for any $x \in \mathscr{L}^2$. In other words, we need to prove that

$$||x\phi||_2 = \sqrt{\int_X (x\phi)^2} = \sqrt{\int_X x^2\phi^2} < \infty$$

Since the function ϕ is bounded by some finite B we get that

$$\phi^2 \leq B^2 \Rightarrow \int_X f_i^2\phi^2 \leq \int_X B^2 f_i^2$$

and therefore

$$||f_i\phi||_2 \leq B\sqrt{\int_X f_i^2} = B||f_i||_2 < \infty$$

□

Definition 5. *Define a new inner product*

$$\langle A, B\rangle_\phi = \langle AM_\phi, B\rangle_{HS}$$

where ϕ is a positive bounded function, or equivalently a positive $\mathscr{L}^\infty$ function.
Note. *This is indeed an inner product because it fulfills all of the following defining properties:*

- *conjugate symmetry:*

$$\begin{aligned}
\langle A, B\rangle_\phi &= \langle AM_\phi, B\rangle_{HS} \\
&= \sum_{j=1}^{n} \langle \phi f_i, g_i\rangle \\
&= \sum_{j=1}^{n} \langle \phi g_i, f_i\rangle \\
&= \langle BM_\phi, A\rangle_{HS} \\
&= \langle B, A\rangle_\phi
\end{aligned}$$

- *linearity*

$$\langle A, B\rangle_\phi + \langle C, B\rangle_\phi = \langle AM_\phi, B\rangle_{HS} + CM_\phi, B\rangle_{HS} = \langle (A + C)M_\phi, B\rangle_{HS} = \langle A + C, B\rangle_\phi$$

$$\langle \alpha A, B\rangle_\phi = \langle \alpha AM_\phi, B\rangle_{HS} = \alpha\langle AM_\phi, B\rangle_{HS} = \alpha\langle A, B\rangle_\phi$$

- *positive-definiteness*

$$\langle A, A\rangle_\phi = \langle AM_\phi, A\rangle_{HS} = \sum_{i=1}^{n} \langle f_i\phi, f_i\rangle = \sum_{i=1}^{n} \int f_i^2\phi$$

  which is greater or equal than 0 since ϕ is positive. Equality happens iff $A = 0$.

Definition 6. *The distance function derived from the newly defined inner product is*

$$||A - B||_\phi = \sqrt{\langle A - B, A - B\rangle_\phi}$$

3 A new metric

3.1 Function alignment

It would be tempting to use the new distance function as our new metric, but we need to take into account the case where $g_i(x) = f_i(x+\alpha)$. In this case, we would like to normalize f_i and g_i in such a way that the distance proposed will minimize the difference between the two.

We therefore introduce a generalized time-shift (or lag) operator T_α to normalize the functions we are trying to compare.

Definition 7. *A translation operators is defined as*

$$(T_h f)(x) = f(h^{-1}(x))$$

where h is a bijection.

Definition 8. *A time-shift translation operator T_β from $\mathscr{L}^2$ to $\mathscr{L}^2$ is defined as a translation operator where the bijective function $h = x - \beta$, so that $h^{-1} = x + \beta$ and*

$$(T_\beta f)(x) = f(x+\beta)$$

Note. *In the special case that f is periodic with the period equal to $\sup(X) - \inf(X)$ we have that*

$$||T_\beta|| = \sup \sqrt{\int_X f(x+\beta)^2} = \sup \sqrt{\int_{X-\beta} f(x)^2} = 1$$

Definition 9. *The generalized time-shift operator T_α is an operator from $(\mathscr{L}^2)^n$ to $(\mathscr{L}^2)^n$ such that*

$$T_\alpha A = \begin{pmatrix} -f_1(x+\alpha_1)- \\ -f_2(x+\alpha_2)- \\ \cdots \\ -f_n(x+\alpha_n)- \end{pmatrix}$$

with $a_i \in \mathbb{R}$.

Note. *T_α does not change the norm of A if f_i is periodic with the period equal to $\sup(X) - \inf(X)$:*

$$\begin{aligned} ||T_\alpha A|| &= \sqrt{\sum_{i=1}^{n} \int_X f_i(x+\alpha_i)^2} \\ &= \sqrt{\sum_{i=1}^{n} \int_{X-\alpha_i} f_i(x)^2} \\ &= \sqrt{\sum_{i=1}^{n} \int_X f_i(x)^2} \\ &= ||A|| \end{aligned}$$

If we consider our functions as being stock prices over time, then this property does not hold.

3.2 Going the distance

Definition 10. *Given a positive $\mathscr{L}^\infty$ function ϕ, define the metric*

$$d_\phi(A,B) = \min_{T_\alpha, T_\beta} ||T_\alpha A - T_\beta B||_\phi$$

where T_α and T_β are two shift operators.
Note. *We can further simplify this to*

$$d_\phi(A,B) = \min_{T_\gamma} ||A - T_\gamma B||_\phi$$

where we combine the two shift operators in one such that $\gamma_i = \beta_i - \alpha_i$ if our functions are periodic as described above.

As an ending note, we derive the following observation: in the case that the functions are periodic, the distance equation can also be written as

$$\begin{aligned} ||A - T_\gamma B||_\phi &= \langle A - T_\gamma B, A - T_\gamma B\rangle_\phi \\ &= ||A||_\phi^2 + ||B||_\phi^2 - 2\langle A, T_\gamma B\rangle_\phi \end{aligned}$$

so the problem of finding T_γ to minimize $d_\phi(A,B)$ becomes the problem of finding T_γ to maximize $\langle A, T_\gamma B\rangle_\phi$.

If the two functions are differentiable, then the problem can be simplified to finding the minimum of

$$h(\gamma) = \langle A, T_\gamma B\rangle_\phi = \int_X A M_\phi T_\gamma B$$

by analyzing h'.

Of course, in the case where the functions are stock prices modeled by Brownian motion, then the functions are nowhere differentiable and the problem becomes non-trivial.

In the discrete case the solution to this problem is well-known if we weaken the condition that T_γ is a translation to T_γ is a orthonormal operator. Yu, Houle and Mio [1] use this result in their work to find a suitable orthonormal transformation that best aligns two shapes.
Lemma 6. *If A and B are $n \times m$ real-valued matrices, then the orthonormal operator P that minimizes the inner product*

$$\langle A, PB\rangle$$

is of the form $P = V_1 V_2^T$ where V_1 and V_2 are the orthogonal matrices from the SVD of $AB^T = V_1^T D V_2$. $\langle\cdot\rangle$ here represents the Frobenius inner product, which is the same as the Hilbert-Schmidt inner product on a discrete space.

Proof.

$$\langle A, PB\rangle = tr(A(PB)^T) = tr(AB^T P^T)$$

Let the SVD of AB^T be $V_1^T D V_2$ where V_1 and V_2 are orthogonal matrices and D is a diagonal matrix. Then

$$\langle A, PB\rangle = tr(AB^T P^T) = tr(V_1^T D V_2 P^T)$$

We know that orthogonal transformations preserve the inner product, so therefore

$$\langle A, PB\rangle = tr(V_1^T D V_2 P^T) = tr(D V_2 P^T V_1^T)$$

Let $K = V_2 P^T V_1^T$. We then have to maximize the expression

$$\langle A, PB\rangle = tr(DK)$$

Let $D = diag(\lambda_1, \lambda_2, \ldots \lambda_n)$ and

$$K = \begin{pmatrix} \cdots \\ -v_i- \\ \cdots \end{pmatrix}$$

Then

$$\langle A, PB\rangle = tr(DK) = \lambda_1 v_{11} + \lambda_2 v_{22} + \ldots \lambda_n v_{nn}$$

But since K is orthogonal, the magnitude of each of the vectors v_i is at most 1! Then

$$\langle A, PB\rangle = tr(DK) = \lambda_1 v_{11} + \lambda_2 v_{22} + \ldots \lambda_n v_{nn} \leq \lambda_1 + \lambda_2 + \ldots \lambda_n$$

with equality when $v_1 = e_1, v_2 = e_2, \ldots v_n = e_n$. Then the maximizing choice of K is $K = I$! Then the maximizing choice of P is

$$V_2 P^T V_1^T = I \iff P = V_2^T V_1$$

$\square$

Note. *In our setting the same solution could be followed to reach a similar result, but since the SVD of AB^T is $V_1^* D V V_2$, where D is a multiplication operator and V_1 and V_2 are partial isometries, there is the problem of conciling T_γ to best approximate $V_2^* V_1$.*

References

[1] Yu Fan, Houle, D. and Mio, W., Dept. of Math., Florida State Univ., Tallahassee, FL, USA; *Learning Metrics for Shape Classification and Discrimination*, 2010