Complex Correntropy Induced Metric Applied to Compressive Sensing with Complex-Valued Data

João P. F. Guimarães*, Aluisio I. R. Fontes*, Felipe B. da Silva‡, Allan de M. Martins†, and Ricardo von Borries‡
* *Federal Institute of Rio Grande do Norte, 1559 Av. Salgado Filho, Natal, RN 59015-000, Brazil*
† *Federal University of Rio Grande do Norte, Campus Universitario, Natal, RN 59078-970, Brazil*
‡ *Department of ECE, University of Texas at El Paso, 500 West University Avenue, El Paso, TX 79968*

***Abstract*—The correntropy induced metric (CIM) is a well-defined metric induced by the correntropy function and has been applied to different problems in signal processing and machine learning, but CIM was limited to the case of real-valued data. This paper extends the CIM to the case of complex-valued data, denoted by Complex Correntropy Induced Metric (CCIM). The new metric preserves the well known benefits of extracting high order statistical information from correntropy, but now dealing with complex-valued data. As an example, the paper shows the CCIM applied in the approximation of ℓ_0-minimization in the reconstruction of complex-valued sparse signals in a compressive sensing problem formulation. A mathematical proof is presented as well as simulation results that indicate the viability of the proposed new metric.**

***Keywords*-Approximation to ℓ_0, complex correntropy induced metric, complex-valued data, compressive sensing.**

I. Introduction

Correntropy is a similarity measure that is capable of extracting high-order statistical information from data [1]. Because of that, it was used to define a robust metric called Correntropy Induced Metric (CIM), which was used in applications in image processing [2] and neural networks [3].

In [4], the authors show that CIM can also be used to estimate ℓ_0 and reconstruct sparse signals in a compressive sensing formulation. It was shown that, by using the CIM, it is possible to reconstruct the sparse signal using fewer measurements when compared with the traditional approaches such as the ℓ_1–norm [4].

Recently, complex correntropy was defined, which makes possible to use correntropy in problems involving complex-valued data [5]. Then, it is now possible to use this new similarity function to update CIM in order to deal with applications involving complex-valued data.

This paper defines a new metric called Complex Correntropy Induced Metric (CCIM). It extends the robustness of complex correntropy to its respective metric. In addition, this paper shows that CCIM is a valid metric and works as an approximation to ℓ_0. We use this metric in compressive sensing problem formulation to recover sparse complex-valued signals. The results from the simulations indicate that CCIM can be used to reconstruct complex-valued signals.

This paper is organized as follows: Section II reviews the correntropy function, its induced metric, and the recently defined complex correntropy. Section III makes a brief presentation of the compressive sensing problem. Section IV defines and proves that the CCIM is a valid metric. Section V contains simulation results and details on how to use the CCIM as an approximation to ℓ_0. The conclusion, Section VI, summarizes the main findings of this paper. Finally, Appendix A contains the proof of the CCIM as an approximation to ℓ_0.

II. Correntropy

Correntropy is defined as [1]

$$V_\sigma(X,Y) = E[K_\sigma(X,Y)] \tag{1}$$

where $K_\sigma(.)$ is any positive-definite kernel with size σ, for $X, Y \in \mathbb{R}^L$, and $E[\,.\,]$ is the expected value operator. Correntropy is a similarity measure because its value is always positive and maximum when vectors X and Y are equal [6]. One can estimate correntropy by using the Gaussian kernel in the estimator

$$\hat{V}_\sigma(X,Y) = \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{L}\sum_{i=1}^{L}\exp\left(-\frac{(x_i-y_i)^2}{2\sigma^2}\right), \tag{2}$$

where σ represents the kernel size.

A. Correntropy Induced Metric

For any two random vectors $X = (x_1, x_2, \ldots, x_L)$ and $Y = (y_1, y_2, \ldots, y_L)$ the CIM, ψ, is defined as [1]

$$\psi(X,Y) = \sqrt{V_\sigma(0,0) - V_\sigma(X,Y)}.$$

B. CIM as an Approximation to ℓ_0

In [4], the authors show that the CIM can be used as an approximation to ℓ_0 of a sparse signal X

$$\|X\|_0 \approx \psi(X,0) = \frac{G_\sigma(0)}{L}\sum_{i=1}^{L}\left(1-\exp\left(-\frac{x_i^2}{2\sigma^2}\right)\right),$$

where $G_\sigma(0) = (\sqrt{2\pi}\sigma)^{-1}$, in a compressive sensing problem formulation.

C. Complex Correntropy

Recently, the correntropy function was defined for the case of complex-valued data [5]

$$V^c_\sigma(Q,W) = E[K_\sigma(Q,W)]$$

where $K_\sigma(.)$ is any positive-definite kernel with kernel size σ, for $Q, W \in \mathbb{C}^L$, and $E[\,.\,]$ is the expected value operator.

By using the Gaussian kernel, correntropy $V_\sigma(X,Y)$ is related to the probability density function of the event $X = Y$ on the real-valued case [7]. In [5], the authors show that one can use the complex Gaussian kernel given by

$$G^c_\sigma(q-w) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(q-w)(q-w)^*}{2\sigma^2}\right), \tag{3}$$

in order to preserve important properties of the real-valued correntropy, such as its probabilistic meaning.

It is possible to estimate the complex correntropy $V^c_\sigma(Q,W)$ using the Gaussian kernel

$$\hat{V}^c_\sigma(Q,W) = \frac{1}{2\pi\sigma^2}\frac{1}{L}\sum_{i=1}^{L}\exp\left(-\frac{(q_i-w_i)(q_i-w_i)^*}{2\sigma^2}\right). \tag{4}$$

III. Compressive Sensing

An S-sparse discrete-time signal $\mathbf{z}$ with length L and S non-zero elements, where $S \ll L$, can be reconstructed from M measurements $\mathbf{y} = \mathbf{A}\,\mathbf{z}$, where $\mathbf{A}$ is called the measurement matrix with size $M \times L$ and $M < L$. It is possible to reconstruct the original signal $\mathbf{z}$ by solving the minimization problem [8]

$$\mathbf{z} = \operatorname{argmin}\|\mathbf{z}\|_0 \text{ such that } \mathbf{y} = \mathbf{A}\mathbf{z} \tag{5}$$

where $\|\mathbf{z}\|_0$ corresponds to the number of nonzero elements of the signal $\mathbf{z}$ i.e.

$$\|\mathbf{z}\|_0 = card\,\{i : z_i \neq 0\}\,. \tag{6}$$

The problem in (5) is NP-Hard and non-convex, which is non ideal [9]. There are many ways to address this problem [9]. The authors in [4] showed that the CIM could be used as an approximation to $\|\mathbf{z}\|_0$ in the case of a real-valued sparse signal $\mathbf{x}$. In this paper, we extend this problem to a complex-valued sparse signal using the complex correntropy [5].

IV. Complex Correntropy Induced Metric

For any two complex-valued random vectors $Q = (q_1, q_2, \ldots, q_L)$ and $W = (w_1, w_2, \ldots, w_L)$, let us define the complex correntropy induced metric (CCIM), ψ^c, as

$$\psi^c(Q,W) = \sqrt{V^c_\sigma(0,0) - V^c_\sigma(Q,W)}. \tag{7}$$

For the CCIM to be a valid metric, it needs to satisfy the following properties: non-negativity, identity, symmetry, and triangle inequality. Most of the proofs are straightforward from the correntropy properties [5].

A. Symmetry & Non-negativity

The proof of these properties follows from the definitions of positive definiteness and symmetry required by the complex correntropy kernel. Since $V^c_\sigma(Q,W) = V^c_\sigma(W,Q)$ holds, then $\psi(Q,W) = \psi(W,Q)$. Furthermore, for the Gaussian kernel case, one could write $0 \leq V^c_\sigma(Q,W) \leq 1/(2\pi\sigma^2)$. Then, the CCIM is both symmetric and non-negative.

B. Identity

Using the Gaussian kernel, if $Q = W$, then $V^c_\sigma(Q,W) = 1/(2\pi\sigma^2)$, which implies $\psi^c(Q,W) = 0$ for $Q = W$.

C. Triangle Inequality

Similarly to the proof for the real-valued case in [6], one can use a kernel mapping and a vector construction in a feature space, which is a well-defined Hilbert space, to show that

$$\psi^c(Q,T) \leq \psi^c(Q,W) + \psi^c(W,T). \tag{8}$$

From the two vectors Q and W, one can construct two new vectors $\tilde{Q} = [\Phi(q_1), \Phi(q_2), \ldots, \Phi(q_L)]$ and $\tilde{W} = [\Phi(w_1), \Phi(w_2), \ldots, \Phi(w_L)]$ in a Hilbert space, where $\Phi(\,.\,)$ is a nonlinear mapping from the input space to a RKHS [6]. Then, the Euclidean distance $ED(\tilde{Q},\tilde{W})$ is

$$\begin{aligned} ED(\tilde{Q},\tilde{W}) &= \sqrt{\langle \tilde{Q}-\tilde{W}, \tilde{Q}-\tilde{W}\rangle} \\ &= \sqrt{\langle \tilde{Q},\tilde{Q}\rangle - 2\langle \tilde{Q},\tilde{W}\rangle + \langle \tilde{W},\tilde{W}\rangle} \end{aligned} \tag{9}$$

or

$$ED(\tilde{Q},\tilde{W}) = \sqrt{2L(G^c_\sigma(0) - V^c_\sigma(Q,W))} = \sqrt{2L}\,\psi^c(Q,W). \tag{10}$$

Therefore,

$$\psi^c(Q,T) = \frac{ED(\tilde{Q},\tilde{T})}{\sqrt{2L}} \leq \frac{ED(\tilde{Q},\tilde{W})}{\sqrt{2L}} + \frac{ED(\tilde{W},\tilde{T})}{\sqrt{2L}} \tag{11}$$

or

$$\psi^c(Q,T) \leq \psi^c(Q,W) + \psi^c(W,T), \tag{12}$$

which completes the proof.

D. Using CCIM as an Approximation to ℓ_0

By considering $Z = X + jS$ and following the concept used in [4] for the real-valued case, it follows

$$\|Z\|_0 \approx \psi^c(Z,0) = \sqrt{\frac{G^c_\sigma(0)}{L}\sum_{i=1}^{L}\left\{1 - \exp\left(-\frac{x_i^2+s_i^2}{2\sigma^2}\right)\right\}}. \tag{13}$$

A complete proof of this approximation is shown in the Appendix (A).

V. Results

This paper follows the minimization strategy shown in [4], which uses the CIM as an approximation to the ℓ_0-minimization of a real-valued signal in a compressive sensing problem. The compressive sensing problem using the CCIM is then formulated as

$$\mathbf{z} = \operatorname{argmin} \psi^c(\mathbf{z}, \mathbf{0}) \text{ such that } \mathbf{y} = \mathbf{A}\mathbf{z}. \tag{14}$$

Note that, as in [4], one can simplify (13) by dropping the square root. The minimization process followed was made by a constrained gradient projection method [10]. The gradient vector $\mathbf{g}$ is computed and then projected onto the null space of $\mathbf{A}^T$. But, since (13) is not analytical in the complex domain, the standard differentiation cannot be applied. One way to address this problem is to use the Wirtinger Calculus as in [5], which is based on the duality between spaces $\mathbb{C}$ and $\mathbb{R}^2$. For more details, see [11]. So, by using the Wirtinger Calculus, one can obtain the derivative

$$[\mathbf{g}]_{L\times 1} = \left[\frac{\partial \psi^c(\mathbf{z},0)}{\partial z_1^*}, \cdots, \frac{\partial \psi^c(\mathbf{z},0)}{\partial z_L^*}\right], \tag{15}$$

where

$$\frac{\partial \psi^c(\mathbf{z},0)}{\partial z_i^*} = \frac{G_\sigma^c(0) z_i}{2L\sigma^2} \exp\left(-\frac{z_i z_i^*}{2\sigma^2}\right). \tag{16}$$

Then, the projection is obtained by

$$\tilde{\mathbf{g}} = \left[\mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\right]\mathbf{g}, \tag{17}$$

so that, to update $\mathbf{z}$, one can write

$$\mathbf{z}_{t+1} = \mathbf{z}_t - \eta \frac{\tilde{\mathbf{g}}}{\|\tilde{\mathbf{g}}\|_2}. \tag{18}$$

A small σ is required to obtain a good approximation to ℓ_0 (see Appendix A). This causes the gradient descent method to be unstable due to the local minimum produced by the small σ. To overcome this problem, [4] has used a kernel annealing technique proposed by [12]. The same procedure was used in this paper with the same parameters as in [4] but with a fixed learning rate $\eta = 0.05$.

In the simulations, the length of the sparse signal is $L = 512$. Four different values of non-zero elements were used, $S = 16, 32, 64, 128$ and their positions were selected randomly with uniform probability. The values of the non-zero elements, both real and imaginary parts, were generated from a standard Gaussian distribution and the number of measurements taken, M, was adjusted from 40 to 300 with step-size of 4. For each value of M, 10^3 Monte Carlo trials were used. The average of these experiments is shown in Fig. 1. The successful reconstruction corresponds to $\|\mathbf{z}-\mathbf{z}_r\|_2 < 10^{-3}$. As in [4], the measurement matrix $\mathbf{A}$ was generated using a Rademacher distribution, and the gradient descent method started with the minimum energy solution of the linear constraint $\mathbf{z}_0 = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{y}$.

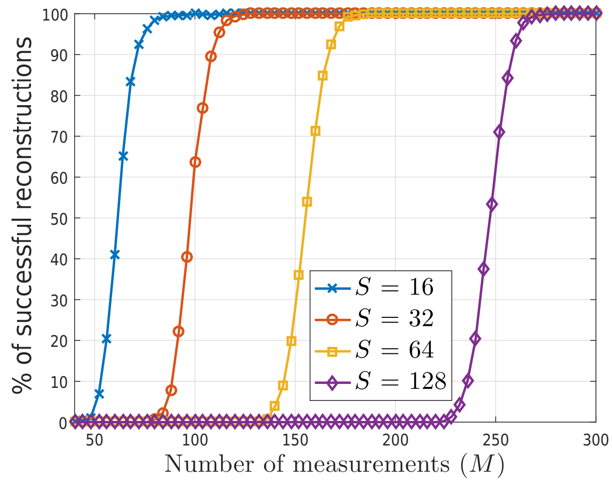

Figure 1. Percentage of successful reconstruction of complex-valued sparse signals with length $L = 512$ and S non-zero elements as a function of the number of measurements (M) taken.

Numerical simulations indicate that it is possible to use the CCIM to approximate ℓ_0 in the reconstruction of sparse complex-valued signals. As expected, an increase in the number of non-zero elements (S) requires an increased number of measurements (M) for successful reconstruction, as in the real-valued case.

VI. Conclusion

This paper defines a new metric for complex-valued random vectors called Complex Correntropy Induced Metric (CCIM). This metric was used to provide a new solution to reconstruct complex-valued sparse signals in a compressive sensing problem formulation. A mathematical proof of the CCIM as an ℓ_0 approximation was presented and tested in simulations with different numbers of measurements in a noiseless setting. The CCIM makes possible to apply the robustness of correntropy into a variety of applications covered by the CIM, but now extended to complex-valued signals. Future work aims into investigating the use of the CCIM in problems with noise.

References

[1] W. Liu, P. P. Pokharel, and J. C. Principe, “Correntropy: Properties and applications in non-gaussian signal processing,” *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, Nov 2007.

[2] X. X. Li, D. Q. Dai, X. F. Zhang, and C. X. Ren, “Structured sparse error coding for face recognition with occlusion,” *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1889–1900, May 2013.

[3] N. Masuyama and C. K. Loo, “Growing neural gas with correntropy induced metric,” in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, Dec 2016, pp. 1–7.

[4] S. Seth and J. C. Principe, “Compressed signal reconstruction using the correntropy induced metric,” in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2008, pp. 3845–3848.

[5] J. P. F. Guimarães, A. I. R. Fontes, J. B. A. Rego, A. de M. Martins, and J. C. Príncipe, “Complex correntropy: Probabilistic interpretation and application to complex-valued data,” *IEEE Signal Processing Letters*, vol. 24, no. 1, 2017.

[6] J. C. Principe, *Information Theoretic Learning: Renyi’s Entropy and Kernel Perspectives*, 1st ed. Springer Publishing Company, Incorporated, 2010.

[7] W. Liu, P. P. Pokharel, and J. C. Principe, “Correntropy: A localized similarity measure,” in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, 2006.

[8] E. J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb 2006.

[9] E. J. Candes and M. B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[10] R. Horie and E. Aiyoshi, “Variable metric gradient projection method and replicator equation,” in *Systems, Man, and Cybernetics, 1999. IEEE SMC ’99 Conference Proceedings. 1999 IEEE International Conference on*, vol. 3, 1999, pp. 515–520 vol.3.

[11] P. Bouboulis and S. Theodoridis, “Extension of wirtinger’s calculus to reproducing kernel hilbert spaces and the complex kernel lms,” *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 964–978, March 2011.

[12] D. Erdogmus, “Information theoretic learning: Renyi’s entropy and its application to adaptive system training,” Ph.D. dissertation, Univ. of Florida, 2002.

Appendix A. Proof of CCIM as an ℓ_0 Approximation

Let $\mathbf{t}$ be the signal that would be obtained by the optimization process when ℓ_0 is minimized, while $\mathbf{r}$ is the signal achieved by the minimization process but using the CCIM instead. Then, one could say that

$$\psi^c(\mathbf{r},0) \leq \psi^c(\mathbf{t},0) \tag{19}$$

and using (7), one could rewrite (19) as

$$\sum_{i=1}^{L} \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) \geq \sum_{i=1}^{L} \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right). \tag{20}$$

The summations in (20) could be written as

$$\sum_{i=1}^{L} \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) = \sum_{\substack{i=1,\\ r_i r_i^*=0}}^{L} \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) + \sum_{\substack{i=1,\\ r_i r_i^*\neq 0}}^{L} \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) \tag{21}$$

and

$$\sum_{i=1}^{L} \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right) = \sum_{\substack{i=1,\\ t_i t_i^*=0}}^{L} \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right) + \sum_{\substack{i=1,\\ t_i t_i^*\neq 0}}^{L} \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right). \tag{22}$$

One could rewrite the summations in (21) and (22) corresponding to $r_i\, r_i^* = 0$ and $t_i\, t_i^* = 0$, respectively, as

$$\sum_{\substack{i=1,\\ r_i r_i^*=0}}^{L} \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) = \sum_{\substack{i=1,\\ r_i r_i^*=0}}^{L} \exp(0) = L - \|\mathbf{r}\|_0 \tag{23}$$

and

$$\sum_{\substack{i=1,\\ t_i t_i^*=0}}^{L} \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right) = \sum_{\substack{i=1,\\ t_i t_i^*=0}}^{L} \exp(0) = L - \|\mathbf{t}\|_0. \tag{24}$$

Substituting (23) and (24) in (21) and (22), respectively, one could write

$$\sum_{i=1}^{L} \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) = L - \|\mathbf{r}\|_0 + \sum_{\substack{i=1,\\ r_i r_i^*\neq 0}}^{L} \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) \tag{25}$$

and

$$\sum_{i=1}^{L} \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right) = L - \|\mathbf{t}\|_0 + \sum_{\substack{i=1,\\ t_i t_i^*=0}}^{L} \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right). \tag{26}$$

Using (25) and (26), it is possible to write (20) as

$$\|\mathbf{r}\|_0 - \sum_{\substack{i=1\\ r_i r_i^*\neq 0}}^{L} \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) \leq \|\mathbf{t}\|_0 - \sum_{\substack{i=1\\ t_i t_i^*\neq 0}}^{L} \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right)$$

or

$$\|\mathbf{r}\|_0 - \|\mathbf{t}\|_0 \leq \sum_{\substack{i=1\\ r_i r_i^*\neq 0}}^{L} \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) - \sum_{\substack{i=1\\ t_i t_i^*\neq 0}}^{L} \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right), \tag{27}$$

and using the fact that $\|\mathbf{r}\|_0 \geq \|\mathbf{t}\|_0$, one can choose σ small enough to make the right-hand side of (27) arbitrarily close to 0^+. Thus

$$\|\mathbf{r}\|_0 \approx \|\mathbf{t}\|_0,$$

which completes the proof.