Generalized Correntropy Matching Pursuit: A novel, robust algorithm for sparse decomposition

Carlos A. Loza, Jose C. Principe

***Abstract*—We introduce a novel variation on the well-known Matching Pursuit (MP) algorithm. In particular, the sparse approximation problem is solved in a greedy scheme using estimated higher-order statistics as similarity measures instead of the somehow limited second-order statistics that perform optimally only under Gaussian assumptions. This is conveyed via the generalized correntropy (GC) function instead of the cross-correlation approach usually utilized in stochastic random processes applications. Additionally, extra flexibility is achieved by the GC parameters that control the behavior of the induced metric. The result is the robust Generalized Correntropy Matching Pursuit (GCMP) algorithm. Furthermore, we present results on two different frameworks dealing with detection and sparse approximation and highlight the robustness of this method in the presence of high-tailed impulsive noise.**

***Index Terms*—Generalized Correntropy, Impulsive Noise, Matching Pursuit, Sparse Decomposition**

I. Introduction

Sparse representations have found a wide range of applications in the machine learning, data mining, and signal processing communities. In particular, sparse coding, compression algorithms, denoising and optimization constitute some of the most relevant areas of study that integrate sparsity as a key framework feature [1]–[3]. In particular for stochastic random processes, sparse decomposition strives to find a parsimonious representation of the time series given that appropriate bases are provided. However most of the times, due to partial knowledge of the underlaying generative system or the presence of noise, this task is reframed as sparse approximation.

Matching Pursuit [4] provides a fast and efficient local solution to this problem by iteratively selecting the highest-correlated atom in an overcomplete dictionary. In addition, the signal representation is not constrained by the Heisenberg uncertainty principle or the wavelet limitation regarding narrow high frequency support. Thus, it has been widely applied to analyze non-stationary time series [5]–[7]. Event though MP is robust against additive white Gaussian noise (AWGN), there is no literature at the time of this paper regarding MP under non-white noise environments. Hence, we propose correntropy [8] as an alternative similarity measure that exploits higher-order statistics of the data. In this way, robustness against high-tailed impulsive noise can be incorporated to the sparse approximation framework.

Carlos A. Loza and Jose C. Principe are with the Computational NeuroEngineering Laboratory (CNEL), University of Florida, Gainesville, FL (email: cloza@ufl.edu, principe@cnel.ufl.edu)

This work was supported by Michael J. Fox Grant 9558

In particular, the generalized correntropy function utilizes the Generalized Gaussian density (GGD) and presents additional shape parameters that expand even further the range of possible induced metrics correntropy possesses. Consequently, GC not only generalizes correntropy, but it also generalizes second-order statics metrics, such as cross-correlation, by behaving like L_2 norm when the parameters are properly set; in this way, one of the lower bounds of GC is the already well utilized MP.

The rest of the paper is organized as follows: Section 2 introduces the generalized correntropy function while Section 3 details the Matching Pursuit implementation using the GGD as cost function. Section 4 presents results on two different types of data and applications, and finally Section 5 concludes the paper with a brief summary and further research directions.

II. Generalized Correntropy

Correntropy was introduced by Liu, Pokharel, and Principe as an attempt to overcome the inherent limitations when working under second-order statistics and Gaussian conditions. Specifically, cross correntropy, or simply correntropy, is a generalized similarity measure between two random variables X and Y:

$$V_\gamma(X, Y) = \mathbf{E}[\kappa_\gamma(X - Y)] \tag{1}$$

where $\kappa_\gamma(X-Y)$ represents a kernel operator with parameter vector γ. Moreover, if the kernel satisfies Mercer's Theorem [9], it induces a nonlinear mapping from the input space to an infinite dimensional reproducing kernel Hilbert space (RKHS). In this case, it has been proved that correntropy is a second-order statistic of the mapped feature space data, while, at the same time, incorporates higher-order moments of the random variable $E = X - Y$ by varying the values of the kernel parameter vector. This allows to define independence of random variables in a more strict manner, i.e. uncorrelatedness in the feature space translates to independence in the input space. Thanks to these attractive properties, correntropy has become a widespread tool in the signal processing and machine learning realms, e.g. classification, clustering, adaptive signal processing, non-linearity tests, non-linear dimensionality reduction, robust filtering, and estimation theory [10]–[18] to name a few.

Most of the aforementioned applications have utilized the Gaussian kernel as the preferred mapping function mostly due to its positive semidefinite property, compliance with Mercer's Theorem, and computational tractability. However, it is possible to incorporate additional degrees of freedom and expand the range of possible properties of the correntropy function. One alternative proposed by Chen et al. [19] is the Generalized Gaussian density, defined by [20] as:

$$G_{\alpha,\beta}(e) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left|\frac{e}{\beta}\right|^{\alpha}\right) \quad (2)$$

where $\Gamma(.)$ is the gamma function, $\alpha > 0$ is the shape parameter, and $\beta > 0$ is the scale (bandwidth) parameter, i.e. for this case the kernel parameter γ is a two-dimensional vector, instead of a unidimensional value, as in the case of the regular Gaussian distribution. Moreover, special cases of the GGD include the Laplace ($\alpha = 1$) and Gaussian ($\alpha = 2$) densities, and a Uniform distribution over $(-\beta, \beta)$ when $\alpha \to \infty$. In practice, in order to incorporate the GGD into the correntropy framework, it is necessary to estimate the true mapping via averages of the available samples of the random variables X and Y, i.e. $\{(x_i, y_i)_{i=1}^{N}\}$. Hence, the sample estimator of the now known as generalized correntropy is

$$\hat{V}_{\alpha,\beta}(X,Y) = \frac{1}{N}\sum_{i=1}^{i=N} G_{\alpha,\beta}(x_i - y_i) \quad (3)$$

Finally, it is worth mentioning that with the additional degrees of freedom GGD provides, it is feasible for the generalized correntropy induced metric (GCIM) to behave like different norms depending on the region, i.e. from L_∞ to L_0, while, on the other hand, the metric induced by a Gaussian-based correntropy mapping only spans from L_2 to L_0.

III. Generalized Correntropy Matching Pursuit

The sparse approximation problem can be posed as an constrained optimization, i.e. in general, for vectors in a finite-dimensional Hilbert space C^Λ:

$$\min_{\Lambda\subset\Omega}\min_{\mathbf{b}\in C^\Lambda}\left\|\mathbf{x} - \sum_{\lambda\in\Lambda}\mathbf{b}_\lambda\varphi_\lambda\right\|_2 \text{ subject to} |\Lambda| \leq m \quad (4)$$

where $\mathbf{x}$ is the input vector, $D = \{\varphi_\omega : \omega \in \Omega\}$ is the dictionary matrix where each one of the its columns is known as an atom; lastly, $\mathbf{b}$ is a list of complex-valued coefficients. If the outermost constraint is ignored, the inner minimization is a least square problem where well-known methods can be utilized; the outer minimization, however, is combinatorial and intractable solutions may surface. This is also categorized as an NP-hard scenario. Possible solutions involve L_1 relaxations such as the Basis Pursuit algorithm [21], L_2 norm use instead of the original L_0 condition which gives raise to the Method of Frames (MOF) [22] and locally optimal algorithms, also known as greedy, that preserve the L_0 norm-based sparsity constraint. From this last category, Matching Pursuit is the most prominent alternative due to its fast implementation and exceptional resolution in both temporal and spectral domains when applied to time series.

Particularly, MP finds the maximum-correlated atom at each iteration, then, linearly subtracts its contribution and repeats the process with the current residue. The decomposition stops after a particular criterion has been met, for instance L runs of the algorithm. Ideally, the reconstructed signal monotonically converges to the original vector in a mean-squared sense, i.e. the reconstruction error power is a non-decreasing function of the number of iterations. Additionally, if the similarity measure is assessed by inner products and the random variables are embedded in time, i.e. stochastic random processes, it is possible to modify the initial cost function in order to accommodate the temporal structure of the data by performing cross correlations instead of inner products. This allows to efficiently compute the decomposition amplitudes and timings via Fast Fourier Transform (FFT) routines. MP has been thoroughly studied and extended by the scientific community; among the most notable variations, we find Orthogonal Matching Pursuit [23], CoSaMP [24], and Kernel Matching Pursuit [25] to name a few.

Nevertheless when working with time series, MP and its extensions usually assume AWGN as the prevalent, most common case even though regular metrics based on second-order statistics, such as inner product or cross correlation, can only perform optimally under Gaussian conditions. Furthermore, one of the main features of correntropy is its robustness against heavy-tailed impulsive noises that are commonly associated to large-amplitude outliers. Hence, it is advantageous to combine this attractive property of correntropy with the greedy approach MP has and the extra flexibility present in the GGD; this results in the Generalized Correntropy Matching Pursuit algorithm. Basically, the inner product is replaced by the generalized correntropy function between each one of the dictionary atoms and an M-dimensional time embedded vector, i.e. $\mathbf{x_i} = [x[i-M+1], x[i-M+2], \ldots x[i]]^T$. Likewise MP, the algorithm stops after a particular criterion has been met; throughout this paper, we use a fixed number of iterations for such condition. The details of this novel variation of Matching Pursuit is presented in Algorithm 1.

Algorithm 1 Generalized Correntropy Matching Pursuit

Parameters: α, β, L

$r[n] \leftarrow x[n]$
for $i = 1 \ldots L$ **do**
  for $j = M \ldots N$ **do**
    $b_q[j] = \hat{V}_{\alpha,\beta}(\varphi_{\mathbf{q}}, \mathbf{r_j}) \qquad q = 1, \ldots, P$
  end for
  $p_i \leftarrow \text{argmax}_q \max_n b_q[n]$
  $\tau_i \leftarrow \text{argmax}_n b_{p_i}[n]$
  $\psi_i \leftarrow b_{p_i}[\tau_i]$
  $r[n] \leftarrow r[n] - \sum_{u=-\infty}^{u=\infty} \psi_i \delta[n - \tau_i - u]\varphi_{p_i}[u]$
end for

IV. Results

A. Detection on Synthetic Data

The first set of results focus on robust detection of events and their corresponding timings in a synthetic time series corrupted by noise. Specifically, a discrete time signal consists of scaled versions of non-overlapping atoms from a 50-dimensional Discrete Cosine Transform (DCT) dictionary; moreover, there is no repetition of bases, the scaling is limited between 0 and 1, and the total length of the time series is 5000 samples. Then, noise with the following distribution is added:

$$p_Z(z) = q\mathbf{N}(0, 0.1) + (1-q)\mathbf{N}(a, 0.1) + (1-q)\mathbf{N}(-a, 0.1) \quad (5)$$

This pdf describes what is known as high-tailed impulsive noise where high-amplitude events occur at low rates; thus, they characterize outliers. Fig. 1 illustrates an example of the time series for $q = 0.9$ and $a = 100$. For this case, we assume we have access to the true dictionary and utilize it for the analysis part. Moreover in order to assess proper detection of the scaled atoms over time, we compute the true positive rate and compare the average performance of Matching Pursuit and Generalized Correntropy Matching Pursuit over a wide range of SNR values. We also know the ground truth here in terms of number of relevant events present in each time series, i.e. 50; thus, we ran the sparse decomposition algorithms for 50 iterations only. Finally for this first experiment, we set $\alpha = 2$ and $\beta = \sqrt{2}\sigma$, i.e. Gaussian distribution similarity measure.

Fig. 2 depicts the average results over 100 simulation trials per SNR value. It is evident that smaller kernel widths exhibit the best consistent behavior; this was expected due to the L_0 norm-like behavior the correntropy induced metric (CIM) possesses for such kernel parameters. In addition, it is remarkable how the lower performance bound for this correntropy-based detector is the regular matching pursuit curve. This constitutes an appropriate proof of concept that reaffirms the fact that correntropy displays characteristics comparable to standard metrics.Finally, as the SNR increases, all the variations of GCMP and MP converge to the same true positive rate due to the relative reduction of high-amplitude outliers.

B. Sparse Decomposition Under Impulsive Noise

For the next experiment, we used a voice recording of the word "gabor" sampled at 1 KHz. The goal is to solve the sparse approximation problem under no knowledge of the true generating dictionary in addition to the presence of strong outliers that resemble skips in the voice recording. For the analysis dictionary, we utilized a 100-dimensional DCT dictionary that provides appropriate temporal and spectral resolution. Once again, impulsive noise is added according to equation 5; for this case $q = 0.99$ and $a = 1000000$ in order to emphasize the outliers components of the noise.

Next, we generated 100 different realizations of the noise process, added it to the voice recording and varied the SNR. MP and GCMP is ran for 75 iterative trials and the normalized power of the reconstructed signal is considered as performance metric. Fig. 3 shows one set of sample results.

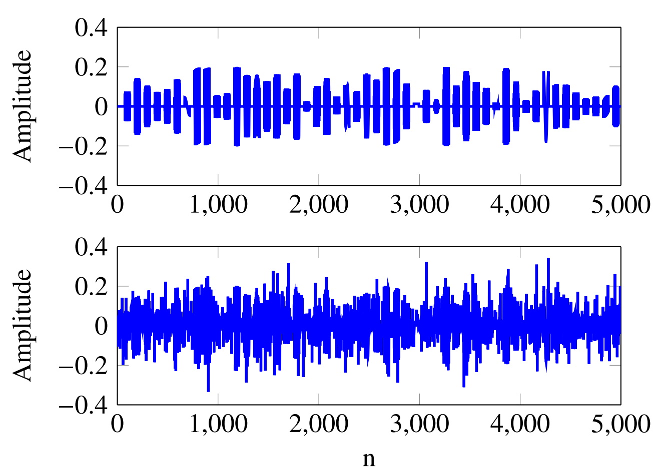

Fig. 1: Upper plot: Sample clean signal over time. Lower plot: Sample signal plus impulsive noise (SNR=2 dB)

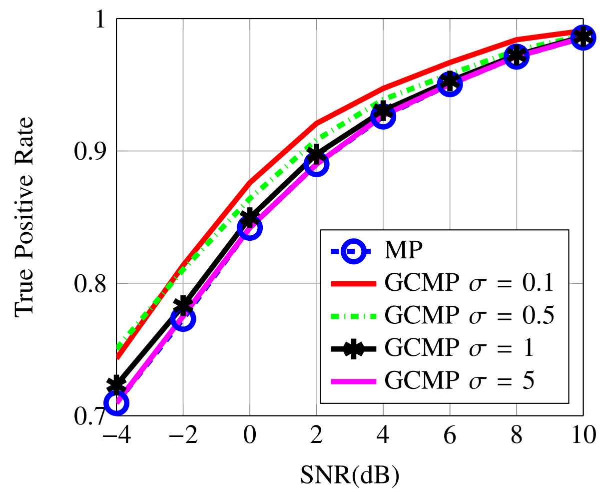

Fig. 2: Average true positive rate for MP and GCMP for $\alpha = 2$ and different kernel width values. The standard deviations are not shown due to comparable range over methods and for visual purposes.

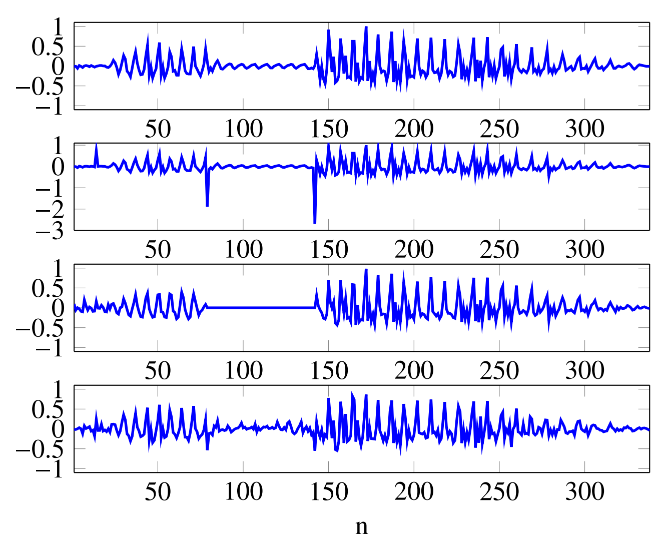

Fig. 3: Example of sparse approximation of the word "gabor" under strong outlier noise. From top to bottom: Original clean signal. Signal plus impulsive noise (SNR=2 dB). Reconstructed signal using MP. Reconstructed signal using GCMP

Particularly, Matching Pursuit completely ignores part of the vowel "a" due to the presence of impulsive noise, i.e. it assumes there is no underlaying relevant events on that block of data. While on the other hand, Generalized Correntropy Matching Pursuit preserves that snippet and successfully eliminates the outlier noise components. Furthermore, Fig. 4 depicts the average results for a fixed $\alpha = 1.5$ and a range of β parameters over different SNR values. GCMP consistently outperforms MP for $\beta = 1$; however, smaller kernel sizes seem to perform suboptimally. This phenomenon has to be further studied; however it might have to do with the interplay between the GDD parameters and the fact that for $\alpha < 2$, the generalized correntropy function will not default to a L_2 metric behavior. Additionally, Table I summarizes the average results for SNR= 2 dB and several combinations of the GDD parameters. It is worth noting that for cases when $\alpha < 2$, the normalized power monotonically increases until achieving the best results at high kernel widths. On the other hand when $\alpha \geqslant 2$, there is clear maximum around $\beta = 1$; after that point, the curve decreases and even reaches the lower bound of MP when $\alpha = 2$. Hence, we are able to demonstrate that GCMP is effectively able to recover the original signal under impulsive noise scenarios.

TABLE I: Average normalized reconstructed signal power of the word "gabor" under impulsive noise. SNR = 2 dB (MP result for this case = 0.93)

α \ β	0.25	0.5	0.75	1	5
0.5	0.65	0.70	0.73	0.75	0.79
1.0	0.83	0.86	0.88	0.89	0.94
1.5	0.78	0.82	0.91	0.97	0.99
2.0	0.71	0.82	0.98	0.96	0.93
2.5	0.64	0.83	0.91	0.91	0.90

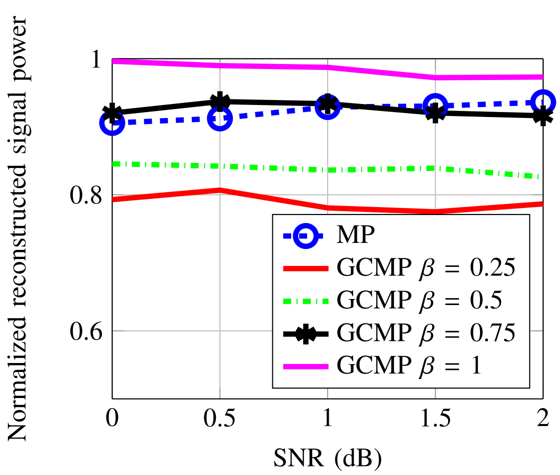

Fig. 4: Average normalized reconstructed signal power of the word "gabor" under impulsive noise. MP and GCMP results for $\alpha = 1.5$ and several β values. The standard deviations are not shown due to comparable range over methods and for visual purposes.

V. Conclusion

The Generalized Correntropy Matching Pursuit allows to incorporate the Generalized Gaussian Distribution as the similarity measure for an iterative, local, greedy solution of the sparse approximation problem. Experimental simulations on detection open the possibility to utilize this technique for channel access methods such as CDMA and even the pervasive GPS system. In addition, the algorithm is able to efficiently recover signals under strong impulsive noise giving as result a sparse representation with remarkable temporal and spectral resolutions.

References

[1] R. G. Baraniuk, "Compressive sensing," *IEEE signal processing magazine*, vol. 24, no. 4, 2007.

[2] J.-L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *Image Processing, IEEE Transactions on*, vol. 11, no. 6, pp. 670–684, 2002.

[3] P. Bühlmann and S. Van De Geer, *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.

[4] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Transactions on*, vol. 41, no. 12, pp. 3397–3415, 1993.

[5] P. Durka and K. Blinowska, "Analysis of eeg transients by means of matching pursuit," *Annals of biomedical engineering*, vol. 23, no. 5, pp. 608–611, 1995.

[6] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *Communications, IEEE Transactions on*, vol. 50, no. 3, pp. 374–377, 2002.

[7] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *Signal Processing, IEEE Transactions on*, vol. 51, no. 1, pp. 101–111, 2003.

[8] W. Liu, P. P. Pokharel, and J. C. Príncipe, "Correntropy: properties and applications in non-gaussian signal processing," *Signal Processing, IEEE Transactions on*, vol. 55, no. 11, pp. 5286–5298, 2007.

[9] V. Vapnik, *The nature of statistical learning theory*. Springer Science & Business Media, 2013.

[10] J.-W. Xu, P. P. Pokharel, A. R. Paiva, and J. C. Príncipe, "Nonlinear component analysis based on correntropy," in *Neural Networks, 2006. IJCNN'06. International Joint Conference on*. IEEE, 2006, pp. 1851–1855.

[11] K.-H. Jeong, W. Liu, S. Han, E. Hasanbelliu, and J. C. Principe, "The correntropy mace filter," *Pattern Recognition*, vol. 42, no. 5, pp. 871–885, 2009.

[12] A. Gunduz and J. C. Principe, "Correntropy as a novel measure for nonlinearity tests," *Signal Processing*, vol. 89, no. 1, pp. 14–23, 2009.

[13] A. Singh and J. C. Principe, "Using correntropy as a cost function in linear adaptive filters," in *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*. IEEE, 2009, pp. 2950–2955.

[14] R. He, W. S. Zheng, and B. G. Hu, "Maximum correntropy criterion for robust face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 8, pp. 1561–1576, 2011.

[15] S. Zhao, B. Chen, and J. C. Principe, "Kernel adaptive filtering with maximum correntropy criterion," in *Neural Networks (IJCNN), The 2011 International Joint Conference on*. IEEE, 2011, pp. 2012–2017.

[16] R. He, B.-G. Hu, W.-S. Zheng, and X.-W. Kong, "Robust principal component analysis based on maximum correntropy criterion," *Image Processing, IEEE Transactions on*, vol. 20, no. 6, pp. 1485–1494, 2011.

[17] B. Chen and J. C. Príncipe, "Maximum correntropy estimation is a smoothed map estimation," *Signal Processing Letters, IEEE*, vol. 19, no. 8, pp. 491–494, 2012.

[18] J. J.-Y. Wang, X. Wang, and X. Gao, "Non-negative matrix factorization by maximizing correntropy for cancer clustering," *BMC bioinformatics*, vol. 14, no. 1, p. 107, 2013.

[19] B. Chen, L. Xing, H. Zhao, N. Zheng, and J. C. Príncipe, "Generalized correntropy for robust adaptive filtering," *arXiv preprint arXiv:1504.02931*, 2015.

[20] M. K. Varanasi and B. Aazhang, "Parametric generalized gaussian density estimation," *The Journal of the Acoustical Society of America*, vol. 86, no. 4, pp. 1404–1415, 1989.

[21] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1998.

[22] I. Daubechies, "Time-frequency localization operators: a geometric phase space approach," *Information Theory, IEEE Transactions on*, vol. 34, no. 4, pp. 605–612, 1988.

[23] J. Tropp *et al.*, "Greed is good: Algorithmic results for sparse approximation," *Information Theory, IEEE Transactions on*, vol. 50, no. 10, pp. 2231–2242, 2004.

[24] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[25] P. Vincent and Y. Bengio, "Kernel matching pursuit," *Machine Learning*, vol. 48, no. 1-3, pp. 165–187, 2002.