Research notes **February 19, 2015**

Summary on scattering transform

Yuansi Chen, Reza Abbasi Asl

Abstract

This report summarizes the research work related to scattering transform, our progress in this direction and thoughts about future directions. In the first section, we summarize the mathematical foundation of scattering transform and describe the two main properties of scattering transform: local translation invariance and Lipschitz continuity to diffeomorphisms. The second section presents the performance of scattering transform on feature extraction in audio and image classification tasks. Then we explore another related direction on the possibility of doing phase recovery with wavelet transform. In the forth section, we provide some visualization of the scattering transform to better understand the transformations that are computed. Finally, we explore the capability of scattering features in modeling V4 neurons.

1 Mathematical Foundation of Scattering Transform

Scattering transform is first mathematically introduced "Group Invariant Scattering" [11] by Mallet. This author attempts to construct operators on $\mathbf{L}^2(\mathbb{R}^d)$ that are translation invariant and Lipschitz continuous to the action of diffeomorphisms. Scattering propagator is a path ordered product of non-linear and non-commuting operators, each of which computes the modulus of a wavelet transform. A local integration of the scattering propagator defines a windowed scattering transform. This transform is proved to be Lipschitz continuous to the action of $\mathbf{C}^2$ diffeomorphisms. The author also show that the windowed scattering transform converges to the scattering transform as window size increases. Hence the windowed scattering transform shares the local translation invariant property.

The two properties that the author expects to have is the translation-invariance and the Lipschitz continuity to action of diffeomorphisms. They could be formulated as follows: Let $f \in \mathbf{L}^2(\mathbb{R}^d)$ be the signal. Let $L_c f(x) = f(x-c)$ denote the translation of f by c.

- **Translation invariance** An operator Φ from $\mathbf{L}^2(\mathbb{R}^d)$ to a Hilbert space $\mathcal{H}$ is translation-invariant if $\Phi(L_c f) = \Phi(f)$ for all $f \in \mathbf{L}^2(\mathbb{R}^d)$ and $c \in \mathbb{R}^d$.
- **Lipschitz continuity and Stability** For stability, we ask the transform to be non-expansive:

  $$\forall (f,h), \|\Phi(f) - \Phi(h)\| \leq \|f - h\|$$

  Now consider a small perturbation of the function $L_\tau f(x) = f(x - \tau(x))$ for $\tau \in \mathbf{C}^2(\mathbb{R}^d)$. A translation invariant operator Φ is said to be Lipschitz continuous to the actions of $\mathbf{C}^2$ diffeomorphisms if for andy compact $\Omega \subset \mathbb{R}^d$ there exists C such that

  $$\|\Phi(f) - \Phi(L_\tau(f))\| \leq C\|f\|(sup_x(\Delta\tau(x)) + sup_x|H\tau(x)|)$$

Note that the operator that computes the modulus of the Fourier transform is an example of translation invariant operator. However, this operator is Lipschitz continuous to the action of diffeomorphisms. Instabilities to deformation appear at high frequencies. This is wavelet transform becomes useful: high frequency instability could be avoided by grouping frequencies into dyadic packets. But the wavelet transform is not translation invariant. The author propose to construct this locally translation invariant operator with scattering procedure along multiple paths and spatial average.

The formal definition of the scattering transform involves cascading of wavelet transforms. The wavelet transform of a signal x computes a convolution of x with a low-pass filter ϕ and convolutions with all higher-frequency wavelets ψ_λ for $\lambda \in \Lambda$:

$$Wx = (x \star \phi(t), x \star \psi_\lambda(t))_{t\in\mathbb{R},\lambda\in\Lambda}$$

The first-order scattering coefficient are obtained by averaging the wavelet modulus coefficients with ϕ:

$$S_1(x, \lambda_1) = |x \star \psi_{\lambda_1}| \star \phi(t)$$

The complementary high-frequency wavelet coefficients of the information lost by this averaging can also be recovered by a second wavelet modulus transform. These coefficients are then averaged by the lowpass filter ? of size T , which ensures local invariance to time-shifts, as with the first-order coefficients. This defines second-order scattering coefficients:

$$S_2(x, \lambda_1, \lambda_2) = ||x \star \psi_{\lambda_1}| \star \lambda_2| \star \phi(t)$$

Iteratively, we can define the scattering network. The final scattering vector aggregates all scattering coefficients for $0 \leq m \leq l$:

$$Sx = (S_m x)_{0\leq m\leq l}$$

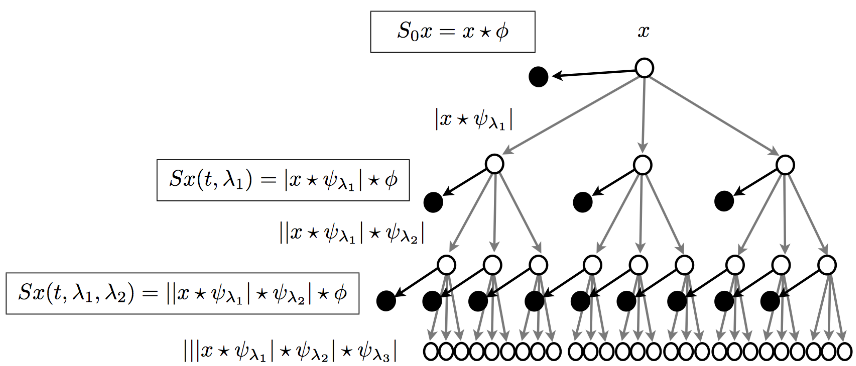

Figure 1: A scattering transform iterates on wavelet modulus operators $|W_m|$ to compute cascades of m wavelet convolutions and moduli stored in $U_m x$, and o output averaged scattering coefficients $S_m x$.

Note that although the author explained the construction from Fourier transform as it is necessary to introduce the wavelet transform, we think that the scattering transform is not the unique operator that satisfies the two main properties. (From frame theory, we think that any frames with bounded conditions on its second and forth moments coupled with local averaging will have these two properties. (To be made precisely, the proof might be nasty.)) We should keep in mind that scattering transform is not the only way to construct locally translation invariant and stable operators.

2 Scattering Transform in Application

2.1 Scattering transform for audio feature extraction

In many audio applications, the first step involves extracting local features from wave files. Mel-frequency cepstrum coefficients (MFCC) is a commonly used feature. It is computed by averaging a spectrogram along the frequency axis according to a mel-frequency scale. This averaging makes the coefficients stable to time-warping deformations. In fact, the MFCC feature is equivalent to two-layer scattering transform with proper choice of wavelets [3]. And higher-layer scattering transform extends MFCC by allowing high-order information. Papers applying scattering transform to phone classification and music genre classification [1, 3, 2] show slightly better performance when compared to standard MFCC features. However, in practice, the scattering transform feature usually occupy more memory than the MFCC features. Later, Sturm [15] evaluated the scattering features on music genre classification with a comparison to sparse representation based approach. It turns out that the sparse representation based approach outperforms the scattering features on this particular task.

We should also note the these scattering transform features are only tested on datasets like TIMIT and GTZAN, where the task is quite simple. We have not found any applications of these features in audio processing pipelines.

2.2 Scattering transform for image feature extraction

To apply scattering transform for image, scattering transform should be extended to use two dimensional wavelets. Scattering transforms build invariant, stable and informative representations through a non-linear, unitary transform, which delocalizes signal information into scattering decomposition paths. They are computed with a cascade of 2D wavelet modulus operators. As the computation of wavelet transform is exactly convolution, a connection to convolutional neural network could also be made [5]. As a result, [5, 7] showed that state-of-the-art classification results could be obtained on hand-written digit recognition and texture classification using scattering transform features.

We have reproduced the results of digit recognition on MNIST using scattering transform features with SVM classifier. Again, we have not found any applications of these features in more complicated vision pipelines. It's not clear whether higher-order scattering transform could provide high level information for image recognition. However, simple naive application of the scattering transform to natural image classification dataset Caltech 101 showed poor performance [13] when compared to traditional computer vision pipeline and also deep convolutional neural network.

2.3 Other work using scattering transform

There are also some recent results on audio texture synthesis with scattering transform in [6]. Chudacek et al. has been used this transform to extract characteristic from fetal heart rate [9]. Poggio et al. also links the scattering transform to their hypothesis about neural tuning.

3 About Inverse Scattering Transform

Scattering coefficients are computed through cascades of wavelet convolutions and modulus operators. For any $m \geq 1$, iterated wavelet modulus convolutions are written:

$$U_m x(t, \lambda_1, ..., \lambda_m) = |||x \star \psi_{\lambda_1}| \star ...| \star \psi_{\lambda_m}(t)|$$

$U_{m+1} = U(U_m)$ with the operator U in the form: $x \to |x \star \psi_\lambda| \lambda \in \Lambda$. Under some particular hypotheses, one can show that U is invertible: the problem is exactly the phase recovery problem with wavelet measurements. The study of inverse of U and of its instabilities is interesting for the comprehension of the scattering operator's behavior. The inverse of scattering transform would guarantee the information is preserved and the stability of the inverse would guarantee the stability of the feature to noise. This constitute another line of research in scattering transform.

3.1 Phase Recovery

The underlying problem about inverse scattering transform is in fact the phase recovery problem under specific measurements. Here we have wavelet functions as measurements.

The phase recovery problem is well studied. Suppose $x \in \mathbb{C}^n$ is a discrete signal and that we are given information about the squared modulus of the inner product between the signal and some vectors z_i, namely,

$$b_i = | < x, z_i > |, i = 1, ..., m.$$

The phase information is lost in these measurements. We still want to recover x from data b, up to a global phase. It is shown that at least one version [14] of the phase recovery problem is NP-hard. Thus, one of the major challenges in the field is to find conditions on m and z_i which guarantee efficient numerical recovery.

Balan, Casazza and Edidin [4] have shown that with probability 1, $4n - 2$ generic measurement vectors (which includes the case of random uniform vectors) suffice for uniqueness in the complex case.

Phase recovery problem is a nonconvex optimization problem. Until recently, this problem was solved commonly using greedy algorithms [10]. This is done by alternating projections on the range of Z and on the nonconvex set of vectors b such that $|b| = |Zx|$. However, these algorithms often stall in local minima. A convex relaxation called PhaseLift was introduced [8] by observing that $|Zx|^2$ is a linear function of $X = xx^*$ which is a rank one Hermitian matrix. The recovery of x is thus expressed as a rank minimization problem over positive semidefinite Hermitian matrices X satisfying some linear conditions. This last problem is approximated by a semidefinite program using trace-norm relaxation. It has been shown that this program recovers x with high probability when A has gaussian independent entries. Based on the same kind of ideas, PhaseCut algorithm [16] was introduced as a convex relaxation to phase recovery problem very similar to the MaxCut problem.

3.2 Phase retrieval with wavelet measurements

Results [4] from frame theory would guarantee that in finite dimensional case, for wavelets with well controlled frame numbers, phases could always be recovered by enough wavelet measurements. However, it is not clear where phase retrieval could be done stably with arbitrary deterministic wavelet measurements. Recent work [12] by Mallet showed that with Cauchy wavelet measurements, phase retrieval could be done uniquely, but the inverse operator is not stable.

4 Visualization of Scattering Transform

We could visualize the wavelets used in scattering transform and optimal stimuli of each output for better understanding of the scattering transform.

4.1 Wavelet filters visualization

Scattering transform on 2D image can be seen as a cascade of Gabor wavelet transforms and modulus operator. In a standard parameter setting, we use wavelets of 8 different orientations and 4 different scales. These 32 wavelet filters are shown here.

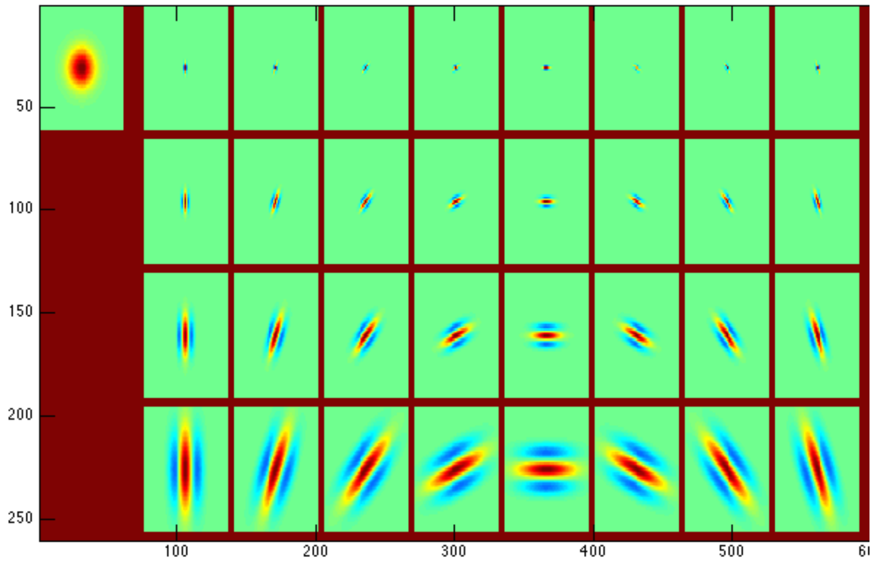

Figure 2: 32 wavelet filters of 8 different orientations and 4 different scales.

4.2 Optimal stimulus visualization

The optimal stimulus of an activation function is defined as the direction that maximum of activation is attained. Given an activation function $f\colon \mathbb{R}^n \to \mathbb{R}$, S unit sphere. The optimal stimulus x^* is

$$x^* = \arg\max_{x \in S} f(x).$$

Projected gradient ascent method could be used to find optimal stimulus. We first compute these optimal stimulus and then visualize it in the image domain. The first layer of scattering transform has 32 outputs. We compute the optimal stimuli and expect it to be of the form of Gabor wavelets. Here we use 10 random initialization for the gradient method. Figure 3 show that in almost all cases, the gradient method converges and the optimal stimulus all look like Gabor wavelets and they are of different scales. This is somewhat expected as the first layer scattering are just combining modulus with wavelet transform.

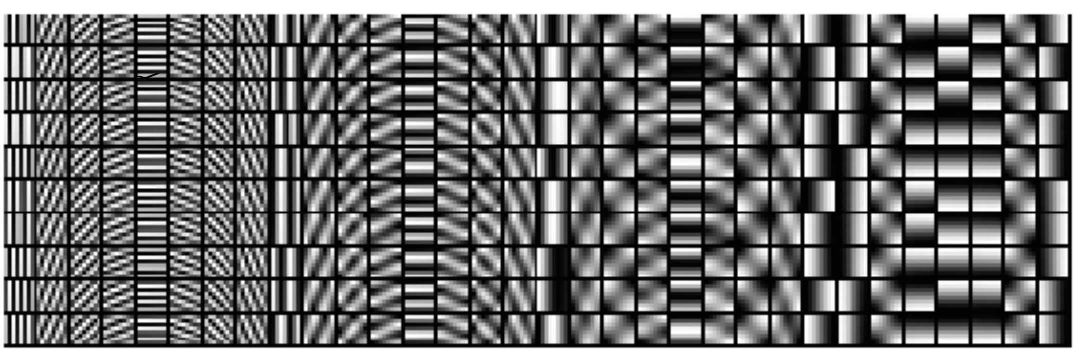

Figure 3: Optimal Stimuli of the First Layer. Each row is one trial of optimal stimuli from 32 wavelets.

The second layer of scattering transform has 384 outputs. It consists of cascading modulus with wavelet transform two times. We compute the optimal stimuli and expect it to be of the composition of two wavelet transforms. Figure 4 show that in each optimal stimulus, there are a high frequency component of one direction and a low frequency component of another direction.

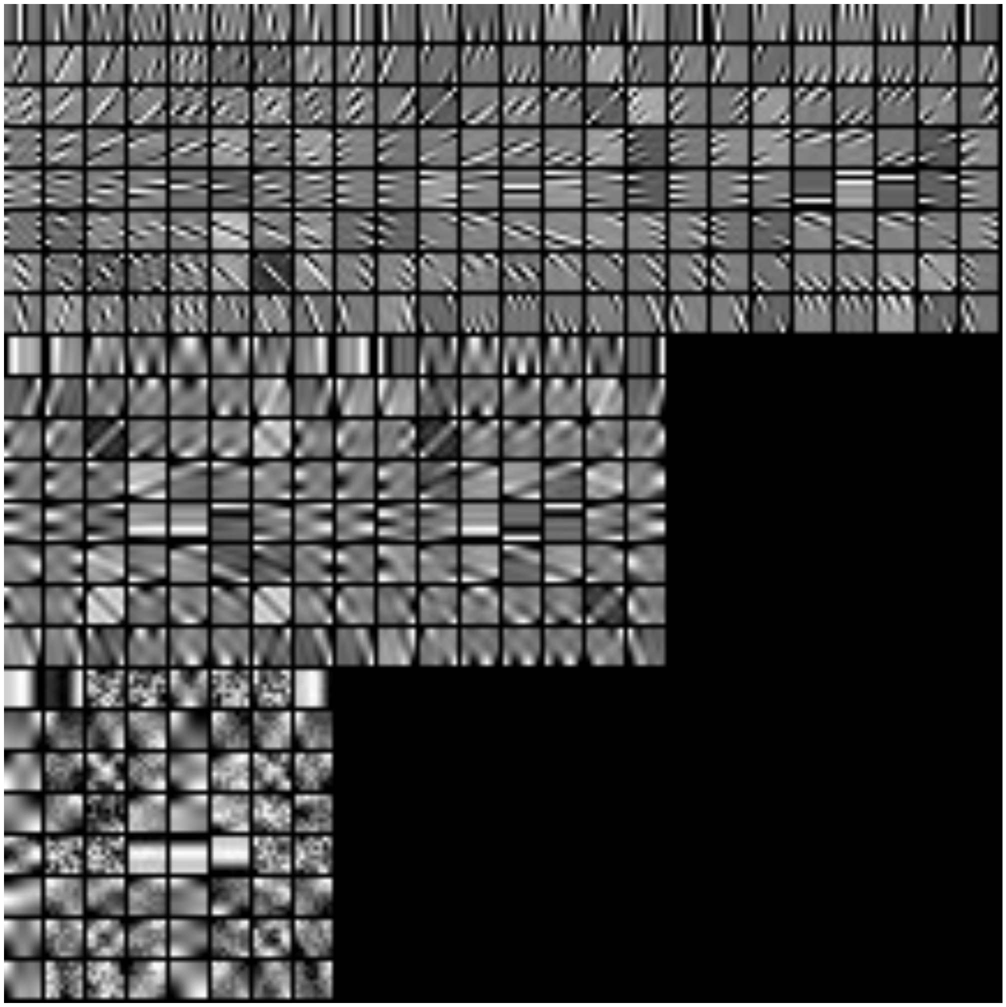

Figure 4: Optimal Stimuli of the Second Layer.

4.3 Most invariant direction

As the scattering transform is claimed to have local translation invariance and stability. We further use second-order Taylor expansion to approximate activation function, in order to determinate the most invariant direction. We denote T the tangent space to S at x^*. By second order Taylor expansion, for h small,

$$f(x^* + h) \approx f(x^*) + g^T h + h^T H h,$$

where g is the gradient and H is the Hessian at x^*. We denote $\tilde{H}$ the projection of H on the tangent space T. By the optimality of x^*, $\tilde{H}$ is a negative semi-definite matrix.

The **most invariant direction** of f at x^* is thus defined as the eigenvector corresponding to the least negative eigenvalue of $\tilde{H}$. In the same way, the least invariant direction of f at x^* is the eigenvector corresponding to the most negative eigenvalue of $\tilde{H}$. Once the most invariant direction v is found, we could visualize the type of invariance. We walk in the direction $\phi(\theta) = \cos(\theta)x^* + \sin(\theta)v$, for small θ, which corresponds to a geodesic on S starting from x^* and moving towards v. (See attached video)

5 Generalized scattering transform

The second layer scattering transform optimal stimulus are already very abundant and redundant. This suggests that statistically the scattering transform computes a variety of statistics which could be potentially useful for any kind of classification tasks. However, based on the visualization we observe that, unlike what we could observe in V4 neuron tuning experiments, cross-like stimulus could not be found among these optimal stimulus of scattering transform. This leads us to introduce the generalized scattering transform.

The idea is to introduce three-dimensional wavelet-like filters in the second layer scattering transforms. These three-dimensional filters would allow for combination of different orientations of wavelets, and thus cross-like stimulus. The third dimension of the filter is built across the different orientations. It is not clear which kind of wavelet transform to use. Further experimentation is needed.

6 Scattering transform on V4 neuron activity prediction

We use two-layer scattering transform as image feature extraction part, followed by sparse coding and linear regression to predict V4 neural activity. Previous model uses hand-crafted wavelet filters as feature extraction part and then sparse coding and linear regression. To have a fair comparison of these two pipelines, we have set the sparse coding and linear regression part to have exactly the same hyper-parameters. A total number of 16 differently oriented wavelets are used in our scattering transform. We use the correlation between the predicted values and the values in testing set as a measure of performance. The following figure shows that the method based on scattering transform outperforms the previous model. Both correlation and smoothed correlation values are calculated. Figure 5 shows the scatter plot of the correlation values in both two correlation measures.

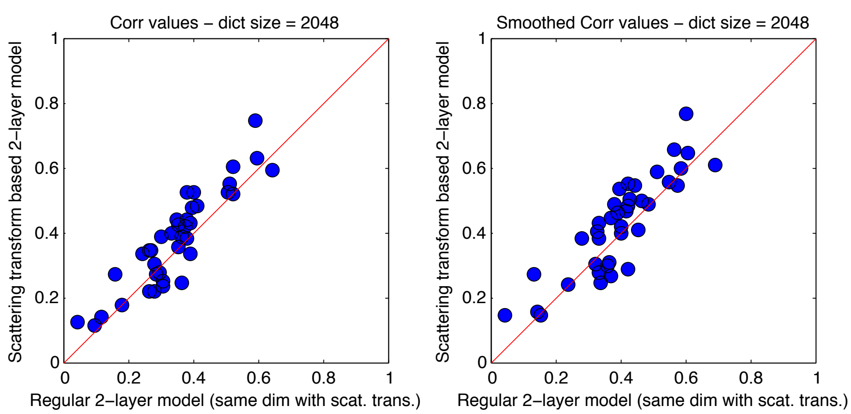

Figure 5: Scatter plot of correlation values. Left: correlation scatter plot comparison between scattering transform and previous model. Right: smoothed correlation scatter plot comparison between scattering transform and previous model.

Note that the two-layer scattering transform pipeline we used here actually have one more layer than the previous model. We have also tried to use two-layer scattering transform without sparse coding layer, the performance is actually worse. Thus it is hard to argue that the scattering transform features are well suited for neuron activity prediction, as the sparse coding layer might play an import role in the process. Further interpretation is needed.

References

[1] Joakim Andén and Stéphane Mallat. Multiscale scattering for audio classification. In *ISMIR*, pages 657–662, 2011.

[2] Joakim Andén and Stéphane Mallat. Scattering representation of modulated sounds. In *Proc. of the 15th Int. Conference on Digital Audio Effects (DAFx-12)*, 2012.

[3] Joakim Andén and Stéphane Mallat. Deep scattering spectrum. *Submitted to IEEE transactions of Signal Processing*, 2013.

[4] Radu Balan, Pete Casazza, and Dan Edidin. On signal reconstruction without noisy phase. *arXiv preprint math/0412411*, 2004.

[5] Joan Bruna and Stéphane Mallat. Classification with scattering operators. *arXiv preprint arXiv:1011.3023*, 2010.

[6] Joan Bruna and Stéphane Mallat. Audio texture synthesis with scattering moments. *arXiv preprint arXiv:1311.0407*, 2013.

[7] Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(8):1872–1886, 2013.

[8] Emmanuel J Candes, Thomas Strohmer, and Vladislav Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.

[9] Václav Chudácek, Joakim Andén, Stéphane Mallat, Patrice Abry, and Muriel Doret. Scattering transform for intrapartum fetal heart rate characterization and acidosis detection. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 2898–2901. IEEE, 2013.

[10] RW Gerchberg. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237, 1972.

[11] Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.

[12] Stéphane Mallat and Irène Waldspurger. Phase retrieval for the cauchy wavelet transform. *arXiv preprint arXiv:1404.1183*, 2014.

[13] Edouard Oyallon, Stéphane Mallat, and Laurent Sifre. Generic deep networks with wavelet scattering. *arXiv preprint arXiv:1312.5940*, 2013.

[14] Haralambos Sahinoglou and Sergio D Cabrera. On phase retrieval of finite-length sequences using the initial time sample. *Circuits and Systems, IEEE Transactions on*, 38(8):954–958, 1991.

[15] Bob L Sturm. Classification accuracy is not enough. *Journal of Intelligent Information Systems*, 41(3):371–406, 2013.

[16] Irène Waldspurger, Alexandre d'Aspremont, and Stéphane Mallat. Phase recovery, maxcut and complex semidefinite programming. *Mathematical Programming*, pages 1–35, 2012.