Global Guarantees for Enforcing Deep Generative Priors by Empirical Risk

Paul Hand HAND@RICE.EDU
Rice University, Department of Computational and Applied Mathematics

Vladislav Voroninski VLAD@HELM.AI
Helm.ai

Editors: Sébastien Bubeck, Vianney Perchet and Philippe Rigollet

Abstract

We examine the theoretical properties of enforcing priors provided by generative deep neural networks via empirical risk minimization. In particular we consider two models, one in which the task is to invert a generative neural network given access to its last layer and another in which the task is to invert a generative neural network given only compressive linear observations of its last layer. We establish that in both cases, in suitable regimes of network layer sizes and a randomness assumption on the network weights, that the non-convex objective function given by empirical risk minimization does not have any spurious stationary points. That is, we establish that with high probability, at any point away from small neighborhoods around two scalar multiples of the desired solution, there is a descent direction. Hence, there are no local minima, saddle points, or other stationary points outside these neighborhoods. These results constitute the first theoretical guarantees which establish the favorable global geometry of these non-convex optimization problems, and they bridge the gap between the empirical success of enforcing deep generative priors and a rigorous understanding of non-linear inverse problems[1].

Keywords: deep learning, generative modeling, nonconvex optimization, compressed sensing

1. Introduction

Exploiting the structure of natural signals and images has proven to be a fruitful endeavor across many domains of science. Breaking with the dogma of the Nyquist sampling theorem, which stems from worst-case analysis, Candès et al. (2006) and Donoho (2006), provided a theory and practice of compressed sensing (CS), which exploits the sparsity of natural signals to design acquisition strategies whose sample complexity is on par with the sparsity level of the signal at hand. On a practical level, compressed sensing has lead to significant reduction in the sample complexity of signal acquisition of natural images, for instance speeding up MRI imaging by a factor of 10. Beyond MRI, compressed sensing has impacted many if not all imaging sciences, by providing a general tool to exploit the parsimony of natural signals to improve acquisition speed, increase SNR and reduce sample complexity. CS has also lead to the development of the fields of matrix completion (Candès and Recht, 2009), phase retrieval, (Candès et al., 2013b,a) and several other subfields (Ahmed et al., 2014) which analogously exploit the sparsity of singular values of low rank matrices as well as sparsity in some basis.

1. Extended abstract. Full version appears as arXiv:1705.07576, v3

Meanwhile, the advent of practical deep learning has significantly improved meaningful compression of images and acoustic signals. For instance, deep learning techniques are now the state of the art across most of computer vision, and have taken the field far beyond where it stood just a few years prior. The success of deep learning ostensibly stems from its ability to exploit the hierarchical nature of images and other signals. There are many techniques and add-on architectural choices associated with deep learning, but many of them are non-essential from a theoretical and, to an extent, practical perspective, with simple convolutional deep nets with Rectified Linear Units (ReLUs) achieving close to the state of the art performance on many tasks (Springenberg et al., 2014). The class of functions represented by such deep networks is readily interpretable as hierarchical compression schemes with exponentially many linear filters, each being a linear combination of filters in earlier layers. Constructing such compression schemes by hand would be quite tedious if not impossible, and the biggest surprise of deep learning is that simple stochastic gradient descent (SGD) allows one to efficiently traverse this class of functions subject to highly non-convex learning objectives. While this latter property has been empirically established in an impressive number of applications, it has so far eluded a completely satisfactory theoretical explanation.

Optimizing over the weights of a neural network or inverting a neural network may both be interpreted as inverse problems (Mallat, 2012). Traditionally, rigorous understanding of inverse problems has been limited to the simpler setting in which the optimization objective is convex. More recently, there has been progress in understanding non-convex optimization objectives for inverse problems, in albeit analytically simpler situations than those involving multilayer neural networks. For instance, Sun et al. (2016) and Bandeira et al. (2016) provide a global analysis of non-convex objectives for phase retrieval and community detection, respectively, ruling out adversarial geometries in these scenarios for the purposes of optimization.

Very recently, deep neural networks have been exploited to construct highly effective natural image priors, by training generative adversarial networks to find a Nash equilibrium of a non-convex game (Goodfellow et al., 2014). The resulting image priors have proven useful in inverting hidden layers of lossy neural networks (Nguyen et al., 2017) and performing super-resolution (Johnson et al., 2016). Naturally, one may ponder whether these generative priors may be leveraged to improve compressive sensing. Indeed, while natural images are sparse in the wavelet basis, a random sparse linear combination of wavelets is far less structured than say a real-world scene or a biological structure, illustrating that a generic sparsity prior is nowhere near tight. The generative priors provided by GANs have already been leveraged to improve compressed sensing in particular domains (Bora et al., 2017). Remarkably, empirical results (Bora et al., 2017) dictate that given a dataset of images from a particular class, one can perform compressed sensing with 10X fewer measurements than what the sparsity prior alone would permit in traditional CS. As GANs and other neural network-based priors improve in modeling more diverse datasets of images, many scenarios in compressed sensing will benefit analogously. Moreover, using generative priors to improve signal recovery in otherwise underdetermined settings is not limited to linear inverse problems, and in principle these benefits should carry over to any inverse problem in imaging science.

In this paper we present the first global analysis of empirical risk minimization for enforcing generative multilayer neural network priors. In particular we show that under suitable randomness assumptions on the weights of a neural network and successively expansive hidden layer sizes, the empirical risk objective for recovering a latent code in $\mathbb{R}^k$ from m linear observations of the last layer of a generative network, where m is proportional to k up to log factors, has no spurious local minima, in that there is a descent direction everywhere except possibly small neighborhoods

around two scalar multiples of the desired solution. Our descent direction analysis is constructive and relies on novel concentration bounds of certain random matrices, uncovering some interesting geometrical properties of the landscapes of empirical risk objective functions for random ReLU'd generative multilayer networks. The tools developed in this paper may be of independent interest, and may in particular lead to global non-asymptotic guarantees regarding convergence of SGD for training deep neural networks.

1.1. Related theoretical work

There has been much recent progress in analysis of non-convex formulations for inverse problems, albeit in analytically simpler situations than those involving multilayer neural networks. For instance, the Sun et al. (2016) and Bandeira et al. (2016) provide a global analysis of non-convex objectives for phase retrieval and community detection, respectively, ruling out adversarial geometries in these scenarios for the purposes of optimization. Additionally, rigorous recovery guarantees for nonconvex formulations exist for phase retrieval (Candes et al., 2015; Chen and Candes, 2015), blind deconvolution (Li et al., 2016; Ma et al., 2017; Huang and Hand, 2017), robust subspace recovery (Maunu et al., 2017), discrete joint alignment (Chen and Candes, 2016), and more.

In related work, the authors of Bora et al. (2017) also study inverting compressive linear observations under generative priors, by proving a restricted eigenvalue condition on the range of the generative neural network. However, they only provide a local guarantee by showing the global minimzier of empirical risk is close to the desired solution. The work does not establish why the global minimum of the nonconvex problem can be reached at all. In addition, Arora et al. (2015) studied inverting neural networks given access to the last layer using an analytical formula that approximates the inverse mapping of a neural network. The results of Arora et al. (2015) are in a setting where the neural net is not generative, and their procedure is at only approximate, and, since it requires observation of the last layer, it is not readily extendable to the compressive linear observation setting. Meanwhile, the optimization problem we study can yield exact recovery, which we observe empirically via gradient descent. Most importantly, in contrast to Bora et al. (2017) and Arora et al. (2015), we provide a global analysis of the non-convex empirical risk objective function and constructively exhibit a descent direction at every point outside a neighborhood of the desired solution and a negative scalar multiple of it. Our guarantees are non-asymptotic, and to the best of our knowledge the first of their kind.

1.2. Main Results

We consider the inverse problem of recovering a vector $y_0 \in \mathbb{R}^n$ from $m \ll n$ linear measurements. To resolve the inherent ambiguity from undersampling, we assume, as a prior, that the vector belongs to the range of a d-layer generative neural network $G : \mathbb{R}^k \to \mathbb{R}^n$, with $k < n$. To recover the vector $y_0 = G(x_0)$, we attempt to find the latent code $x_0 \in \mathbb{R}^k$ corresponding to it. We consider a generative network modeled by $G(x) = \text{relu}(W_d \ldots \text{relu}(W_2 \, \text{relu}(W_1 x_0)) \ldots)$, where $\text{relu}(x) = \max(x, 0)$ applies entrywise, $W_i \in \mathbb{R}^{n_i \times n_{i-1}}$, n_i is the number of neurons in the ith layer, and $k = n_0 < n_1 < \cdots < n_d = n$. We consider linear measurements of $G(x_0)$ given by the

sampling matrix $A \in \mathbb{R}^{m \times n}$ and consider $k < m \ll n$. The problem at hand is:

$$\begin{aligned} \text{Let:} \quad & x_0 \in \mathbb{R}^k, A \in \mathbb{R}^{m \times n}, W_i \in \mathbb{R}^{n_i \times n_{i-1}} \text{ for } i = 1 \dots d, \\ & G(x) = \text{relu}(W_d \dots \text{relu}(W_2 \, \text{relu}(W_1 x_0)) \dots), \\ & y_0 = G(x_0), \\ \text{Given:} \quad & W_1 \dots W_d, A, \text{ and observations } A y_0, \\ \text{Find:} \quad & x_0. \end{aligned}$$

This problem can be viewed in two ways: (1) as above, given compressive measurements of a vector with the prior information that it belongs to the output of a generative neural network, find that vector; or (2), given compressive observations of the output of a generative neural network, find the latent code corresponding to the network's output by inverting the neural network and compression simultaneously.

As a way to solve the above problem, we consider minimizing the empirical risk objective

$$f(x) := \frac{1}{2} \left\| AG(x) - Ay_0 \right\|_2^2. \tag{1}$$

As this objective is nonconvex, there is no *a priori* guarantee of efficiently finding the global minimum (Murty and Kabadi, 1987). Approaches such as gradient descent could in principle get stuck in local minima, instead of finding the desired global minimizer x_0.

In this paper, we consider a fully-connected generative network $G : \mathbb{R}^k \to \mathbb{R}^n$ with Gaussian weights and no bias term, along with a Gaussian sampling matrix $A \in \mathbb{R}^{m \times n}$. We show that under appropriate conditions and with high probability, f has a strict descent direction everywhere outside two small neighborhoods of x_0 and a negative multiple of x_0. We assume that the network is sufficiently *expansive* at each layer, $n_i = \Omega(n_{i-1} \log n_{i-1})$, and that there are a sufficient number of measurements, $m = \Omega(kd \log(n_1 \cdots n_d))$. Let $D_v f(x)$ be the unnormalized one-sided directional derivative of f at x in the direction v: $D_v f(x) = \lim_{t \to 0^+} \frac{f(x+tv) - f(x)}{t}$. Our main result is as follows:

Theorem 1 *Fix $\epsilon > 0$ such that $K_1 d^8 \epsilon^{1/4} \leqslant 1$, and let $d \geqslant 2$. Assume $n_i \geqslant c n_{i-1} \log n_{i-1}$ for all $i = 1 \dots d$ and $m > cdk \log \prod_{i=1}^d n_i$. Assume that for each i, the entires of W_i are i.i.d. $\mathcal{N}(0, 1/n_i)$, and the entries of A are i.i.d. $\mathcal{N}(0, 1/m)$ and independent from $\{W_i\}$. Then, on an event of probability at least $1 - \sum_{i=1}^d \tilde{c} n_i e^{-\gamma n_{i-1}} - \tilde{c} e^{-\gamma m}$, we have the following. For all nonzero x and x_0, there exists $v_{x,x_0} \in \mathbb{R}^k$ such that the one-sided directional derivatives of f satisfy*

$$\begin{aligned} & D_{-v_{x,x_0}} f(x) < 0, \quad && \forall x \notin \mathcal{B}(x_0, K_2 d^3 \epsilon^{1/4} \|x_0\|_2) \cup \mathcal{B}(-\rho_d x_0, K_2 d^{13} \epsilon^{1/4} \|x_0\|_2) \cup \{0\}, \\ & D_v f(0) < 0, && \forall v \neq 0, \end{aligned}$$

where ρ_d is a positive number that converges to 1 as $d \to \infty$. Here, c and γ^{-1} are constants that depend polynomially on ϵ^{-1}, and $\tilde{c}, K_1, K_2$ are universal constants.

In particular, under the assumptions of the theorem, with high probability there are no local optima or critical points outside of the two specified neighborhoods. Also, note that while the weights of any layer of the network are assumed to be i.i.d. Gaussian, there is no assumption on the independence between W_i and W_j for $i \neq j$.

This theorem will be proven by showing the sufficiency of two deterministic conditions on G and A, and then by showing that Gaussian G and A of appropriate sizes satisfy these conditions with the appropriate probability. The first deterministic condition is on the spatial arrangement of the network weights within each layer. We say that the matrix $W \in \mathbb{R}^{n\times k}$ satisfies the *Weight Distribution Condition* with constant ϵ if for all nonzero $x, y \in \mathbb{R}^k$,

$$\left\| \sum_{i=1}^{n} 1_{w_i\cdot x>0} 1_{w_i\cdot y>0} \cdot w_i w_i^t - Q_{x,y} \right\| \leqslant \epsilon, \text{ with } Q_{x,y} = \frac{\pi-\theta_0}{2\pi} I_k + \frac{\sin\theta_0}{2\pi} M_{\hat{x}\leftrightarrow\hat{y}}, \tag{2}$$

where $w_i \in \mathbb{R}^k$ is the ith row of W; $M_{\hat{x}\leftrightarrow\hat{y}} \in \mathbb{R}^{k\times k}$ is the matrix[2] such that $\hat{x} \mapsto \hat{y}$, $\hat{y} \mapsto \hat{x}$, and $z \mapsto 0$ for all $z \in \operatorname{span}(\{x, y\})^\perp$; $\hat{x} = x/\|x\|_2$ and $\hat{y} = y/\|y\|_2$; $\theta_0 = \angle(x, y)$; and 1_S is the indicator function on S. The norm in the left hand side of (2) is the spectral norm. Note that an elementary calculation[3] gives that $Q_{x,y} = \mathbb{E}[\sum_{i=1}^{n} 1_{w_i\cdot x>0} 1_{w_i\cdot y>0} \cdot w_i w_i^t]$ for $w_i \sim \mathcal{N}(0, I_k/n)$. As the rows w_i correspond to the neural network weights of the ith neuron in a layer given by W, the WDC provides a deterministic property under which the set of neuron weights within the layer given by W are distributed approximately like a Gaussian. The WDC could also be interpreted as a deterministic property under which the neuron weights are distributed approximately like a uniform random variable on a sphere of a particular radius. Note that if $x = y$, $Q_{x,y}$ is an isometry up to a factor of $1/2$.

The second deterministic condition is that the compression matrix acts like an isometry on pairs of differences of vectors in the range of $G : \mathbb{R}^k \to \mathbb{R}^n$. We say that the compression matrix $A \in \mathbb{R}^{m\times n}$ satisfies the *Range Restricted Isometry Condition (RRIC)* with respect to G with constant ϵ if for all $x_1, x_2, x_3, x_4 \in \mathbb{R}^k$,

$$\begin{aligned}\Big|\Big\langle A\big(G(x_1) - G(x_2)\big), A\big(G(x_3) - G(x_4)\big)\Big\rangle - \Big\langle G(x_1) - G(x_2), G(x_3) - G(x_4)\Big\rangle\Big| \\ \leqslant \epsilon \|G(x_1) - G(x_2)\|_2 \|G(x_3) - G(x_4)\|_2.\end{aligned} \tag{3}$$

We can now state our main deterministic result.

Theorem 2 *Fix $\epsilon > 0$ such that $K_1 d^8 \epsilon^{1/4} \leqslant 1$, and let $d \geqslant 2$. Suppose that G is such that W_i has the WDC with constant ϵ for all $i = 1 \dots d$. Suppose A satisfies the RRIC with respect to G with constant ϵ. Then, for all nonzero x and x_0, there exists $v_{x,x_0} \in \mathbb{R}^k$ such that the one-sided directional derivatives of f satisfy*

$$\begin{aligned} &D_{-v_{x,x_0}} f(x) < 0, \quad \forall x \notin \mathcal{B}(x_0, K_2 d^3 \epsilon^{1/4}\|x_0\|_2) \cup \mathcal{B}(-\rho_d x_0, K_2 d^{13}\epsilon^{1/4}\|x_0\|_2) \cup \{0\}, \\ &D_y f(0) < 0, \qquad \forall y \neq 0, \end{aligned}$$

where ρ_d is a positive number that converges to 1 as $d \to \infty$, and K_1 and K_2 are universal constants.

2. A formula for $M_{\hat{x}\leftrightarrow\hat{y}}$ is as follows. If $\theta_0 = \angle(\hat{x}, \hat{y}) \in (0, \pi)$ and R is a rotation matrix such that $\hat{x}$ and $\hat{y}$ map to e_1 and $\cos\theta_0 \cdot e_1 + \sin\theta_0 \cdot e_2$ respectively, then $M_{\hat{x}\leftrightarrow\hat{y}} = R^t \begin{pmatrix} \cos\theta_0 & \sin\theta_0 & 0 \\ \sin\theta_0 & -\cos\theta_0 & 0 \\ 0 & 0 & 0_{k-2} \end{pmatrix} R$, where 0_{k-2} is a $k-2 \times k-2$ matrix of zeros. If $\theta_0 = 0$ or π, then $M_{\hat{x}\leftrightarrow\hat{y}} = \hat{x}\hat{x}^t$ or $-\hat{x}\hat{x}^t$, respectively.
3. To do this calculation, take $x = e_1$ and $y = \cos\theta_0 \cdot e_1 + \sin\theta_0 \cdot e_2$ without loss of generality. Then each entry of the matrix can be determined analytically by an integral that factors in polar coordinates.

In the case that $A = I_n$, the RRIC is trivially satisfied, and we get the following corollary about inverting multilayer neural networks.

Corollary 3 (Approximate Invertibility of Multilayer Neural Networks) *If G is a d-layer neural network such that W_i satisfies the WDC with constant ϵ for all $i = 1 \dots d$, then the function $f(x) = \|G(x) - G(x_0)\|_2$ has no stationary points outside of a neighborhood around x_0 and $-\rho_d x_0$.*

In the case of a Gaussian network with Gaussian measurements, the WDC and RRIC are satisfied with high probability if the network is sufficiently expansive and there are a sufficient number of measurements.

Proposition 4 *Fix $0 < \epsilon < 1$. Assume $n_i \geqslant c n_{i-1} \log n_{i-1}$ for all $i = 1 \dots d$ and $m > cdk \log \Pi_{i=1}^{d} n_i$. Assume the entires of W_i are i.i.d. $\mathcal{N}(0, 1/n_i)$, and the entries of A are i.i.d. $\mathcal{N}(0, 1/m)$. Then, W_i satisfies the WDC with constant ϵ for all i and A satisfies the RRIC with respect to G with constant ϵ with probability at least $1 - \sum_{i=1}^{d} \tilde{c} n_i e^{-\gamma n_{i-1}} - \tilde{c} e^{-\gamma m}$. Here, c and γ^{-1} are constants that depend polynomially on ϵ^{-1}, and $\tilde{c}$ is a universal constant.*

As stated after Theorem 1, no assumption is made on the independence between W_i and W_j for $i \neq j$. While Proposition 4 is stated for $A \in \mathbb{R}^{m \times n}$ with i.i.d. Gaussian entries, it also applies in the case of any random matrix that satisfies the following concentration of measure condition:

$$\mathbb{P}\big(|\|Ax\|_2^2 - \|x\|_2^2| \geqslant \epsilon \|x\|_2^2\big) \leqslant 2e^{-m c_0(\epsilon)},$$

for any fixed $x \in \mathbb{R}^n$, where $c_0(\epsilon)$ is a positive constant depending only on ϵ. In particular, Proposition 4 and hence Theorem 1 extends to the case of where the entries of A are independent Bernoulli random variables (and the entries of W_i are Gaussian). See Baraniuk et al. (2008) for more.

Acknowledgments

PH is partially supported by NSF DMS-1464525.

References

Ali Ahmed, Benjamin Recht, and Justin Romberg. Blind deconvolution using convex programming. *IEEE Transactions on Information Theory*, 60(3):1711–1732, 2014.

Sanjeev Arora, Yingyu Liang, and Tengyu Ma. Why are deep nets reversible: A simple theory, with implications for training. *CoRR*, abs/1511.05653, 2015.

Bandeira, Boumal, and Voroninski. On the low-rank approach for semidefinite programs arising in synchronization and community detection. *JMLR*, 49:1–22, 2016.

Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G. Dimakis. Compressed sensing using generative models. *arXiv:1703.03208*, 2017.

E. Candès, Y. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *SIAM J. Imaging Sci.*, 6(1):199–225, 2013a. doi: 10.1137/110848074.

Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717, 2009.

Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006. ISSN 1097-0312. doi: 10.1002/cpa.20124.

Emmanuel J Candès, Thomas Strohmer, and Vladislav Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Comm. Pure Appl. Math.*, 66(8):1241–1274, 2013b.

Emmanuel J Candes, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval via wirtinger flow: Theory and algorithms. *Information Theory, IEEE Transactions on*, 61(4):1985–2007, 2015.

Yuxin Chen and Emmanuel Candes. Solving random quadratic systems of equations is nearly as easy as solving linear systems. In *Advances in Neural Information Processing Systems*, pages 739–747, 2015.

Yuxin Chen and Emmanuel Candes. The projected power method: An efficient algorithm for joint alignment from pairwise differences. *arXiv preprint arXiv:1609.05820*, 2016.

David Donoho. For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59(6), 2006.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

Wen Huang and Paul Hand. Blind deconvolution by a steepest descent algorithm on a quotient manifold. *arXiv preprint arXiv:1710.03309*, 2017.

Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711. Springer, 2016.

Xiaodong Li, Shuyang Ling, Thomas Strohmer, and Ke Wei. Rapid, robust, and reliable blind deconvolution via nonconvex optimization. *arXiv preprint arXiv:1606.04933*, 2016.

Cong Ma, Kaizheng Wang, Yuejie Chi, and Yuxin Chen. Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion and blind deconvolution. *arXiv preprint arXiv:1711.10467*, 2017.

S. Mallat. Group invariant scattering. *Comm. Pure Appl. Math.*, 65(10):13311398, 2012.

Tyler Maunu, Teng Zhang, and Gilad Lerman. A well-tempered landscape for non-convex robust subspace recovery. *arXiv preprint arXiv:1706.03896*, 2017.

Katta G. Murty and Santosh N. Kabadi. Some np-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39(2):117–129, Jun 1987.

Anh Nguyen, Jeff Clune, Yoshua Bengio, Alexey Dosovitskiy, and Jason Yosinski. Plug & play generative networks: Conditional iterative generation of images in latent space. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, pages 3510–3520. IEEE, 2017.

Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin A. Riedmiller. Striving for simplicity: The all convolutional net. *CoRR*, abs/1412.6806, 2014.

Ju Sun, Qing Qu, and John Wright. A geometric analysis of phase retrieval. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pages 2379–2383. IEEE, 2016.