As engineering and the sciences become increasingly data and computation driven, the importance of seeking succinct data representation and developing efficient optimization methods has expanded to touch almost *every* stage of the data analysis pipeline, ranging from signal and data acquisition to modeling and prediction. For many mission critical applications in neuroscience, physics, computational microscopy, and biomedical/hyperspectral imaging, there is a pressing need for effective, guaranteed data representation models and efficient optimization methods to analyze the massive amount of data we created.

While the challenges in computing with physical data are many and varied, basic recurring issues arise from *nonlinearities* at different stages of this pipeline: (*i*) the representation models we built are often highly nonlinear, and (*ii*) the measurements taken from phyiscal sensors are nonlinear. These naturally result in *nonconvex* problems[1], where guaranteeing their correctness (i.e., the global optimality) used to be notoriously difficult. In the worst case, a nonconvex function has (*i*) *spurious* local minima (Figure 1a) and (*ii*) "*flat*" saddle points (Figure 2a), that in theory even finding a local minimizer is NP-hard [17] – let alone the global minima.

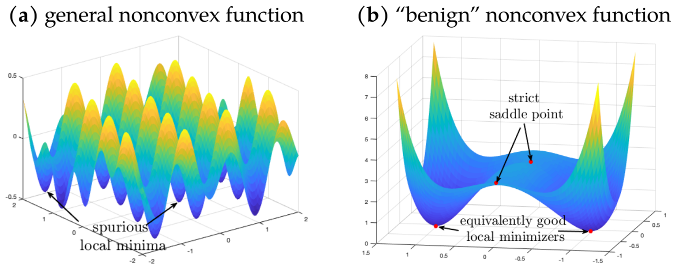

Figure 1: **general vs. benign nonconvex landscape.**

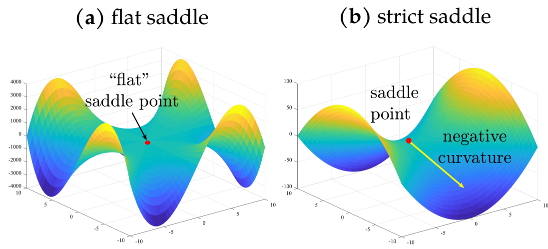

Figure 2: **strict vs. flat saddle points.**

For nonconvex problems in machine learning and data science, my work [1–12] has delivered new and surprising insights into the mysteries of nonconvex optimization: the *symmetry* structures lead to *benign* geometric properties of their optimization landscapes in the sense that

(i) the (only!) local minimizers are *symmetric* versions of the ground truth (Figure 1b);
(ii) for each saddle point there is *negative curvature* in directions that breaks symmetry (Figure 2b),

which *alleviates* the hardness dictated by the worst-case theory (Figures 1a and 2a) and *enables* efficient global optimization. My work has reshaped our understandings of nonconvex problems, leading to more flexible, effective, and guaranteed data representation models, as well as dramatically more efficient and scalable computational tools [4–7, 11, 18, 19]. In the slogan form,

my research develops *efficient* methods and *global* theory for *nonconvex* optimization, producing new computational tools and theoretical guarantees for *representation learning* and *imaging sciences*.

Learning Low-complexity Representations from High-dimensional Data

Thanks to the blessing of dimensionality (e.g., sparsity or low rankness), in practice the *intrinsic* dimension of high dimensional data is often much *lower* than its ambient dimension. While we used to manually *craft* representations to capture the low dimensional information in the past [20], it has been demonstrated that *learned* representations [21] show much *superior* performance in various applications of signal processing, machine learning, theoretical neuroscience, and many other fields [22–24]. However, most of underlying representation models are nonconvex, which raise tremendous challenges in optimization and guaranteeing the correctness.

Learning sparsifying dictionaries [1–3, 9–11]. The goal of dictionary learning (DL) is to produce a sparse model for an observed dataset $\boldsymbol{Y} = [\boldsymbol{y}_1, \dots, \boldsymbol{y}_p] \in \mathbb{R}^{n \times p}$. Namely, we seek matrices $\boldsymbol{A}$ and $\boldsymbol{X}$ such that

$$\underbrace{\boldsymbol{Y}}_{\text{data}} \approx \underbrace{\boldsymbol{A}}_{\text{dictionary}} \underbrace{\boldsymbol{X}}_{\text{sparse coefficients}},$$

with $\boldsymbol{X}$ as *sparse* as possible. Sparsity is desirable for data compression, and to facilitate tasks ranging from low-level image processing to high-level visual recognition [22–24]. This problem exhibits a *signed permutation* (SP) *symmetry*: for a given pair $(\boldsymbol{A}, \boldsymbol{X})$, and any SP matrix $\boldsymbol{\Gamma}$, the pair $(\boldsymbol{A\Gamma}, \boldsymbol{\Gamma}^* \boldsymbol{X})$ approximates $\boldsymbol{Y}$ *equally* well, with the same number of nonzero entries. The SP symmetry *prevents* any natural convex relaxation, while until recently our understandings of existing nonconvex approaches are very limited [25, 26].

[1] It denotes any problem that is not convex [1–15], where convex problems [16] are well-studied with every local minimizer being global.

My geometric analysis [9–11] provides the *first* global theoretical explanations for the DL problem: (*i*) all local minimizers are *global* in the sense that they are all *equivalent* up to a SP symmetry (Figure 3a); (*ii*) any other critical points can be efficiently escaped via negative curvature in symmetry breaking directions (Figure 2b). The new geometric understandings lead to the first efficient and provable polynomial methods of solving DL globally with optimal sparsity level on $\boldsymbol{X}$, and demonstrated on natural images (Figure 4a).

Convolutional sparse coding [3]. Inspired by deconvolutional networks [27], the task of convolutional dictionary learning (CDL) is to learn the filters $\{\boldsymbol{a}_k\}_{k=1}^K$ along with the sparse codes $\{\boldsymbol{x}_{ik}\}_{1\le i\le p, 1\le k\le K}$ from the data $\boldsymbol{y}_i = \sum_{k=1}^K \boldsymbol{a}_k \circledast \boldsymbol{x}_{ik} (1 \le i \le p)$. The CDL model provides more efficient representations of the data [28], and recently draws connection to convolutional neural network [29]. Aside from the SP symmetry, the problem exhibits extra *shift symmetry* in the sense that cyclic shift[2] creates equivalent solution pair $(\mathrm{s}_\ell[\boldsymbol{a}_k], \mathrm{s}_{-\ell}[\boldsymbol{x}_{ik}])$.

By reducing the shift symmetry to SP symmetry via formulating the CDL as an overcomplete DL problem, my work [3] provides the *first* theoretical guarantees for solving CDL. Similar to DL, the discrete symmetry structures lead to benign optimization landscape, that descent methods provably solve CDL with simple initialization.

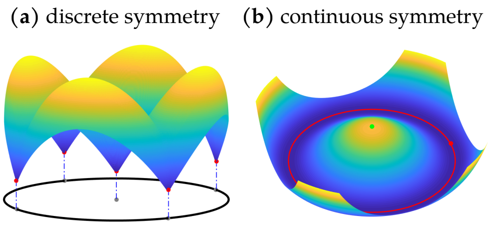

Figure 3: Symmetry leads to benign nonconvex geometry.

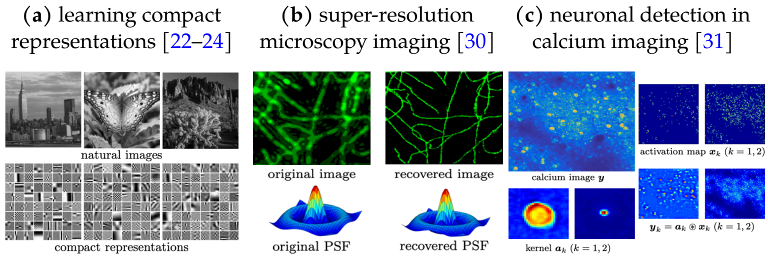

Figure 4: applications in machine learning and imaging science.

Solving Nonlinear Inverse Problems in Imaging Sciences

Inverse problems [32] are *pervasive* in scientific discovery and decision-making for complex, natural, engineered, and societal systems, and thus are of paramount importance across many disciplines including imaging sciences. However, for imaging applications, most of inverse problems are highly *ill-posed*. Although certain low-complexity structure (e.g., sparsity) alleviates the difficulty, the major challenge remains often due to the *nonlinearity* of the measurements and the underlying sensing models, which result in nonconvex optimization problems.

Sparse blind deconvolution [4, 5]. Deconvolving *sparse* point sources from their convolution with an *unknown* point spread function (PSF) finds many applications in neuroscience [31], computational microscopy imaging [30], geophysics [33], astronomy [34], etc. The basic task is to *simultaneously* recover $\boldsymbol{a}$ and $\{\boldsymbol{x}_i\}_{i=1}^p$ from

$$\underbrace{\boldsymbol{y}_i}_{\text{measurement}} = \underbrace{\boldsymbol{a}}_{\text{PSF}} \circledast \underbrace{\boldsymbol{x}_i}_{\text{sparse point sources}}, \quad 1 \le i \le p,$$

where in the general form the problem can be reviewed as a special case of CDL. This problem exhibits an intrinsic *shift symmetry*: for any solution pair $(\boldsymbol{a}, \{\boldsymbol{x}_i\}_{i=1}^p)$ and shift $\mathrm{s}_\ell[\cdot]$ of any length $\ell > 0$, the shifted pair $(\mathrm{s}_\ell[\boldsymbol{a}], \{\mathrm{s}_{-\ell}[\boldsymbol{x}_i]\}_{i=1}^p)$ produces *equivalent* good solutions in the sense that $\boldsymbol{a} \circledast \boldsymbol{x}_i = \mathrm{s}_\ell[\boldsymbol{a}] \circledast \mathrm{s}_{-\ell}[\boldsymbol{x}_i]$.

In the multi-measurement settings $p \ge \mathrm{poly}(n)$, my work [4] shows that shift symmetry introduces benign geometries for nonconvex landscapes (Figure 3a), such that vanilla gradient descent with random initialization finds *exact* solutions with *linear* convergence rate. In the single measurement case ($p = 1$) with *short* PSF $\boldsymbol{a}$ and *long* sequence $\boldsymbol{x}$, similar geometric intuitions helped to develop efficient nonconvex solvers [5], which beat state-of-art optimization methods. The proposed methods have been demonstrated on several imaging applications such as super-resolution microscopy imaging [30] (Figure 4b) and Calcium imaging [31] (Figure 4c).

Phase retrieval [7, 8]. Another important inverse problem in imaging science is, how to retrieve phase information of an unknown complex signal $\boldsymbol{x} \in \mathbb{C}^n$ from *nonlinear* magnitude measurements $\{y_k = |\boldsymbol{a}_k^* \boldsymbol{x}|\}_{k=1}^m$. Finding effective solutions has broad applications in X-ray crystallography, lenseless microscopy, diffraction and array imaging, and optics [35, 36]. The nonlinear measurement induces a *rotational symmetry* (Figure 3b) in the sense that rotating the signal $\boldsymbol{x}$ by an arbitrary phase $\theta \in (0, 2\pi)$ creates another feasible solution $\boldsymbol{x}e^{i\theta}$. Although convex approaches remove this symmetry by lifting the problem into a higher dimension [37], they result in a huge semidefinite programming problem that is *impractical* to solve for most imaging applications.

[2] Here, $\mathrm{s}_\ell[\boldsymbol{v}]$ is a shift operator that cyclic shifts a vector $\boldsymbol{v}$ of length ℓ.

In contrast, my work [8] considers a natural nonconvex least-squares formulation whose problem size is much smaller. When the sensing vectors $\{\boldsymbol{a}_k\}_{k=1}^m$ are *generic* random, my geometric analysis [8] leads to efficient iterative methods recovering the complex signal $\boldsymbol{x}$ up to a rotational symmetry, starting from arbitrary initialization. Again, our success derives from the benign geometric structure induced by the rotational symmetry (Figure 3b): the function has a large sample limit, which (*i*) has no spurious local minima, and (*ii*) can be optimized efficiently. Moreover, based on the geometric intuitions, my subsequent work [7] proves linear convergence of simple gradient descent methods for more *structured* convolutional sensing models.

Future Research Plans

Nonconvex problems with benign symmetry and geometry structures are way beyond what I have studied here – my work has inspired recent study of nonconvex problems such as shallow neural network, low-rank matrix recovery, tensor decomposition, and synchronization, to name a few. [38–40] provide a contemporary overview of recent progress. It is an exciting time to work on nonconvex optimizations ranging from theory to practice.

Towards disciplined nonconvex optimization theory. Despite successes aforementioned, our understandings of nonconvex optimization is still *far* from satisfactory – the analysis is delicate, case-by-case, and pertains to problems with elementary symmetry (e.g., rotation or permutation) and simple manifold (e.g., sphere).

- *A Unified Theory.* Analogous to the study of convex functions [16], there is a pressing need for simpler analytic tools, to identify and generalize benign properties for new nonconvex problems. Our initial investigations [3, 6] show promises to identify general conditions and operations preserving benign geometric structures.
- *Complicated Symmetries and Constraints.* Nonconvex problems in practice often involve either *multiple symmetries* (e.g., Fourier phase retrieval) or *complicated manifolds* (e.g., Stefiel manifold [41]). More technical tools need to be developed towards a better understanding, despite recent endeavors made by me [6] and others [41, 42].
- *Nonsmoothness.* In many scenarios we *inevitably* face nonconvex problems with *nonsmooth* formulations [4, 6, 7], for better promoting solution sparsity or robustness. However, most of our current analysis is *local* [6, 7], and (subgradient) optimization [6] could be slow in convergence. Towards *global* analysis and *fast* optimization, we might need advanced tools from variational analysis [43] and development of efficient 2nd-order methods.

Understandings and improvements of Deep Neural Networks. Empirical success of deep learning is another demonstration of the power of representation learning [21] and nonconvex optimization [1–12]. The lack of solid understanding limits its application to scientific discovery, and many other mission-critical applications.

- *Taming Symmetries.* As aforementioned, solutions to nonlinear inverse problems and learned representations are often expected to be invariant to translation, rotation, and permutation. The relation of naïve input-output pairs of a network could be one-to-many, resulting in poor generalization performance [44]. This appears to be a fundamental problem, which requires innovative training protocol or network architecture designs [45].
- *Understanding Optimization and Generalization.* Mysteries of efficient optimization of highly nonconvex training loss and generalization of over-parameterized network cannot be explained by classical learning theory [46]. Recent advances based on the neural tangent kernel [47] does not yet explain the many successes observed in various challenging machine learning tasks [48]. A potential satisfactory theory could be starting from solving well-defined inverse problems [49]: given a desired accuracy, the theory should predict a valid combination of network architecture, optimization method, and sample complexity for solving the problem.
- *Learning to Optimize.* Learning based optimizations [50] unroll existing iterative algorithms by replacing components with learned operators. They demonstrate superior performance in terms of convergence and model robustness, which could also be a good start for developing more interpretable network architectures.

Nonconvex optimization for Computational Imaging. By integration of sensing systems and computational algorithms, computational imaging is transforming our way of sensing the world. It overcomes hardware limits via solving computational challenges [51]. Microscopy imaging [36], biomedical imaging [52], geophysics [33], optics [35] are some representative examples, whose core computational task involves solving *highly ill-posed, nonlinear inverse problems*. As illustrated in aforementioned examples, most nonlinear inverse problems in imaging contain intrinsic symmetries, so that target solutions are equivalent to a symmetry ambiguity. Leveraging geometric intuitions for nonconvex optimization, my goal is to advance our understandings of these problems, with new problem formulations and improved recovery guarantees. These advances will enable design of more efficient sensors, as well as guaranteed and fast optimization algorithms. I am eager to engage with experts from sensing, imaging, and a wide range of application domains, using nonconvex optimization to solve large-scale scientific problems, and developing software packages with both high efficiency and accuracy as we did in [5].

Acknowledgement and Future Funding Plans

My research has been generously supported by *NSF CCF/IIS*, *Microsoft fellowship*, and *Moore-Sloan fellowship*. Recently, I submitted a funding proposal as a Co-PI to *NSF DMS*. In the future, I believe the scope of my research will be diversely supported from agencies such as *NSF, NIH, ONR, DoD, and DARPA*. If the position was offered, I plan to submit my first proposal to *NSF CCF/CIF small* on nonconvex methods for sparse deconvolution and deep networks. Based on feedback, I will submit my *NSF Career* proposal on nonconvex optimization theory during my second year, and apply to other early career awards such as *AFOSR/ONR YIP, Sloan fellowship*, etc. Furthermore, I will actively seek interdisciplinary collaboration with experts from neuroscience, bioengineering, and physics, to develop open-source optimization software packages leading to breakthroughs in imaging problems of various domain sciences, which could be potentially funded through *NSF IIS/CDS&E*, and *NIH R01*.

References

Part I – My Selected Publications

[1] **Qing Qu** *, Zhihui Zhu *, Xiao Li, M. C. Tsakiris, John Wright, and René Vidal. Finding the sparsest vectors in a subspace: theory, algorithms, and applications. *In Preparation for IEEE Signal Processing Magazine* (***Invited Paper*** *for Special Issue on Nonconvex Optimization*), 2020 [abstract].

[2] Yuqian Zhang, **Qing Qu**, and John Wright. Geometry and symmetry in nonconvex optimization. *In Preparation for Proceedings of IEEE*, 2020.

[3] **Qing Qu**, Yuexiang Zhai, Xiao Li, Yuqian Zhang, and Zhihui Zhu. Geometric analysis of nonconvex optimization landscapes for overcomplete learning. *In Submission to Journal of Machine Learning Research*, 2019 (among **top 20** highest review scores of 2594 submissions to ICLR'20, according to OpenReview.net) [link] [pdf].

[4] **Qing Qu**, Xiao Li, and Zhihui Zhu. Exact recovery of multichannel sparse blind deconvolution via gradient descent. *Under Revision for SIAM Journal on Imaging Sciences*, 2019 (NeurIPS'19 **spotlight top 3%**) [link] [pdf] [code].

[5] Yenson Lau *, **Qing Qu** *, Han-Wen Kuo, Pengcheng Zhou, Yuqian Zhang, and John Wright. Short-and-sparse deconvolution — A geometric approach. *In Submission to IEEE Transactions on Pattern Analysis and Machine Intelligence and ICLR'20*, 2019 [link] [pdf] [code] [website].

[6] Xiao Li *, Shixiang Chen *, Zengde Deng, **Qing Qu**, Zhihui Zhu, and Anthony Man Cho So. Riemannian algorithms for nonsmooth optimization over Stiefel manifold. *In Submission to SIAM Journal on Optimization*, 2019 [link] [pdf].

[7] **Qing Qu**, Yuqian Zhang, Yonina C. Eldar, and John Wright. Convolutional phase retrieval via gradient descent. *Accepted at IEEE Trans. on Information Theory*, 2019 [link] [pdf].

[8] Ju Sun, **Qing Qu**, and John Wright. A geometric analysis of phase retrieval. *Foundations of Computational Mathematics*, 18(5):1131–1198, 2018 [link] [pdf].

[9] **Qing Qu**, Ju Sun, and John Wright. Finding a sparse vector in a subspace: Linear sparsity using alternating directions. *IEEE Trans. on Information Theory* (*short version in NeurIPS'14*), 62(10):5855–5880, 2016 [link] [pdf].

[10] Ju Sun, **Qing Qu**, and John Wright. Complete dictionary recovery over the sphere I: Overview and the geometric picture. *IEEE Trans. on Information Theory* (*short version in ICML'15*), 63(2):853–884, 2016 [link] [pdf].

[11] Ju Sun, **Qing Qu**, and John Wright. Complete dictionary recovery over the sphere II: Recovery by Riemannian trust-region method. *IEEE Trans. on Information Theory*, 63(2):885–914, 2016 [link] [pdf].

[12] Ju Sun, **Qing Qu**, and John Wright. When are nonconvex problems not scary? *NeurIPS Workshop on Nonconvex Optimization for Machine Learning: Theory and Practice*, 2016 [link] [pdf].

[13] **Qing Qu**, Nasser M. Nasrabadi, and Trac D. Tran. Subspace vertex pursuit: A fast and robust near-separable nonnegative matrix factorization method for hyperspectral unmixing. *IEEE Journal of Selected Topics in Signal Processing*, 9(6):1142–1155, 2015 [link] [pdf].

[14] **Qing Qu**, Nasser M. Nasrabadi, and Trac D. Tran. Abundance estimation for bilinear mixture models via joint sparse and low-rank representation. *IEEE Trans. on Geoscience and Remote Sensing*, 52(7):4404–4423, 2013 [link] [pdf].

[15] Xiaoxia Sun, **Qing Qu**, Nasser M. Nasrabadi, and Trac D. Tran. Structured priors for sparse-representation-based hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 11(7):1235–1239, 2013 [link] [pdf].

$*$ denotes equal contributions.

Part II – Other Publications

[16] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

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

[18] Rong Ge, Furong Huang, Chi Jin, and Yang Yuan. Escaping from saddle points—online stochastic gradient for tensor decomposition. In *Conference on Learning Theory*, pages 797–842, 2015.

[19] Nicolas Boumal, Pierre-Antoine Absil, and Coralia Cartis. Global rates of convergence for nonconvex optimization on manifolds. *IMA Journal of Numerical Analysis*, 39(1):1–33, 2018.

[20] Stephane Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 2008.

[21] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: a review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013.

[22] Michael Elad. *Sparse and redundant representations: from theory to applications*. Springer Science & Business Media, 2010.

[23] John Wright, Yi Ma, Julien Mairal, Guillermo Sapiro, Thomas S Huang, and Shuicheng Yan. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 98(6):1031–1044, 2010.

[24] Julien Mairal, Francis Bach, and Jean Ponce. Sparse modeling for image and vision processing. *Foundations and Trends® in Computer Graphics and Vision*, 8(2-3):85–283, 2014.

[25] Karin Schnass. On the identifiability of overcomplete dictionaries via the minimisation principle underlying k-svd. *Applied and Computational Harmonic Analysis*, 37(3):464–491, 2014.

[26] Sanjeev Arora, Rong Ge, Tengyu Ma, and Ankur Moitra. Simple, efficient, and neural algorithms for sparse coding. *Journal of Machine Learning Research*, 40(2015), 2015.

[27] Matthew D Zeiler, Dilip Krishnan, Graham W Taylor, and Rob Fergus. Deconvolutional networks. In *2010 IEEE Computer Society Conference on computer vision and pattern recognition*, pages 2528–2535. IEEE, 2010.

[28] Hilton Bristow, Anders Eriksson, and Simon Lucey. Fast convolutional sparse coding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 391–398, 2013.

[29] Vardan Papyan, Yaniv Romano, Jeremias Sulam, and Michael Elad. Theoretical foundations of deep learning via sparse representations: a multilayer sparse model and its connection to convolutional neural networks. *IEEE Signal Processing Magazine*, 35(4):72–89, 2018.

[30] Bo Huang, Mark Bates, and X. Zhuang. Super-resolution fluorescence microscopy. *Annual review of biochemistry*, 78:993–1016, 2009.

[31] Eftychios A Pnevmatikakis, Daniel Soudry, Y. Gao, T. A Machado, Josh Merel, David Pfau, Thomas Reardon, Yu Mu, Clay Lacefield, Weijian Yang, et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron*, 89(2):285–299, 2016.

[32] Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*, volume 89. SIAM, 2005.

[33] Kjetil F Kaaresen and Tofinn Taxt. Multichannel blind deconvolution of seismic signals. *Geophysics*, 63(6):2093–2107, 1998.

[34] J. Starck, E. Pantin, and F. Murtagh. Deconvolution in astronomy: a review. *Publications of the Astronomical Society of the Pacific*, 114(800):1051, 2002.

[35] Yoav Shechtman, Yonina C Eldar, Oren Cohen, Henry Nicholas Chapman, Jianwei Miao, and Mordechai Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE Signal Processing Magazine*, 32(3):87–109, 2015.

[36] Euan McLeod and Aydogan Ozcan. Unconventional methods of imaging: computational microscopy and compact implementations. *Reports on Progress in Physics*, 79(7):076001, 2016.

[37] E. J Candès, Y. C Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *SIAM Review*, 57(2):225–251, 2015.

[38] Prateek Jain, Purushottam Kar, et al. Non-convex optimization for machine learning. *Foundations and Trends® in Machine Learning*, 10(3-4):142–336, 2017.

[39] Ju Sun. Provable nonconvex methods/algorithms. `https://sunju.org/research/nonconvex/`.

[40] Yuejie Chi, Yue M Lu, and Yuxin Chen. Nonconvex optimization meets low-rank matrix factorization: an overview. *IEEE Transactions on Signal Processing*, 67(20):5239–5269, 2019.

[41] Jiang Hu, Xin Liu, Zaiwen Wen, and Yaxiang Yuan. A brief introduction to manifold optimization, 2019. arXiv: `1906.05450 [math.OC]`.

[42] Yuexiang Zhai, Zitong Yang, Zhenyu Liao, John Wright, and Yi Ma. Complete dictionary learning via l^4-norm maximization over the orthogonal group, 2019. arXiv: `1906.02435 [cs.LG]`.

[43] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[44] Aharon Azulay and Yair Weiss. Why do deep convolutional networks generalize so poorly to small image transformations?, 2018. arXiv: `1805.12177 [cs.CV]`.

[45] T. S Cohen, M. Geiger, J. Köhler, and M. Welling. Spherical CNNs. In *International Conference on Learning Representations*, 2018.

[46] René Vidal, Joan Bruna, Raja Giryes, and Stefano Soatto. Mathematics of deep learning, 2017. arXiv: `1712.04741 [cs.LG]`.

[47] Simon S. Du, Jason D. Lee, Haochuan Li, Liwei Wang, and Xiyu Zhai. Gradient descent finds global minima of deep neural networks, 2018. arXiv: `1811.03804 [cs.LG]`.

[48] L. Chizat, E. Oyallon, and F. Bach. On lazy training in differentiable programming, 2018. arXiv: `1812.07956 [math.OC]`.

[49] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.

[50] Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, pages 399–406. Omnipress, 2010.

[51] Joseph N Mait, Gary W Euliss, and Ravindra A Athale. Computational imaging. *Advances in Optics and Photonics*, 10(2):409–483, 2018.

[52] M Lustig, D. L Donoho, J. M Santos, and J. M Pauly. Compressed sensing MRI. *IEEE Signal Processing Magazine*, 25:72–82, 2008.