Deep Learning Initialized Phase Retrieval

Raunak Manekar Zhong Zhuang Kshitij Tayal Vipin Kumar Ju Sun
Computer Science and Engineering & Electrical and Computer Engineering
University of Minnesota, Twin Cities
{manek009,zhuan143,tayal,kumar001,jusun}@umn.edu

Abstract

Phase retrieval (PR) consists of estimating 2D or 3D objects from their Fourier magnitudes and takes a central place in scientific imaging. At present, most iterative methods for PR work well only when the initialization is close enough to the solution and fail otherwise. But there has been no general way of obtaining desired initialization. In this paper, we show that a carefully designed deep learning pipeline can consistently generate reliable initialization, so that the subsequent iterative methods can solve the PR problem and produce high-quality solutions. Technically, PR is an inverse problem containing three forward symmetries, and naive deployment of end-to-end deep learning for PR yields poor initialization. We explain why the symmetries cause the learning difficulty and propose a novel strategy that substantially improves the estimation. Overall, the proposed method solves PR in regimes not accessible by the previous methods, and our work synergizes deep learning and iterative methods to solve a difficult scientific inverse problem.

1 Introduction

Given the oversampled Fourier magnitudes $\boldsymbol{Y} = |\mathcal{F}(\boldsymbol{X})|^2 \in \mathbb{R}^{m_1 \times m_2}$ of $\boldsymbol{X} \in \mathbb{C}^{n_1 \times n_2}$ ($m_i \geq n_i$ for $i = 1, 2$), is it possible to recover $\boldsymbol{X}$? This is the phase retrieval (PR) problem central to scientific imaging and numerous other fields [1, 2]. In imaging, the complex phases of $\mathcal{F}(\boldsymbol{X})$ are missing because practical detectors cannot record complex phases. As a nonlinear inverse problem, PR is tricky to solve: (1) for any $\boldsymbol{X}$, global phase shift $\boldsymbol{X} \mapsto e^{i\theta}\boldsymbol{X}$ for all $\theta \in (-\pi, \pi]$, overall spatial shift of the nonzero content of $\boldsymbol{X}$, and 2D conjugate flipping of $\boldsymbol{X}$ all leave the measurement $\boldsymbol{Y}$ unchanged (see Fig. 1; [1]). So the best one can hope for is recovery up to these intrinsic symmetries; and (2) the mapping $\boldsymbol{X} \mapsto |\mathcal{F}(\boldsymbol{X})|^2$ is *generically* injective up to the intrinsic symmetries when $m_i \geq 2n_i - 1$ [4] for $i = 1, 2$. So in this paper, we always assume $m_i \geq 2n_i - 1$ for $i = 1, 2$ and solving PR is up to the intrinsic symmetries.

In practice, PR is often solved by iterative methods such as HIO [5], RAAR [6] and difference map [7]; see [8] for an updated review of these methods. However, they work well when support (i.e., locations of nonzero elements) of $\boldsymbol{X}$ is provided with reasonable accuracy. Otherwise, even the most sophisticated iterative methods fail to work in practice. The failure is mostly due to stagnation caused by the intrinsic symmetries [9].

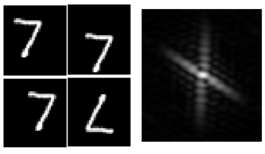

Figure 1: Shift and flipping symmetries in PR. **Left**: shifted and flipped copies of the digit "7"; **Right**: their common Fourier magnitudes. Image credit: [3]

Deep learning has brought about new prospects of solving difficult inverse problems, of which PR is an instance. One can phrase PR as a regularized optimization problem: $\min_{\boldsymbol{X}} \ell(\boldsymbol{Y}, |\mathcal{F}(\boldsymbol{X})|^2) + \lambda\Omega(\boldsymbol{X})$, and then replace ℓ, or Ω, or mappings inside specific numerical methods for PR using data-driven neural network modules. This approach is taken in, e.g., [10, 11], where HIO is needed to produce

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

good initialization—for simpler inverse problems, such special initialization is not required [12]. More radical is the *end-to-end approach*, where a neural network is trained to directly approximate the inverse mapping or its proxies. [13–16] have taken this approach and shown promising results. Here, we take a critical view of the initial successes.

Difficulty of learning with symmetries When solving nonlinear inverse problems with symmetries, the end-to-end approach may face the difficulty of approximating highly oscillatory functions. This issue has recently been elucidated in [17]. We summarize the main argument using the learning square root example: suppose we randomly sample real values x_i's and form a training set $\{x_i, x_i^2\}$ and try to learn the square-root function using the end-to-end approach, allowing both positive and negative outputs. Now if we think of the function determined by the training set, which the neural network is trying to approximate, it is highly oscillatory (see Fig. 2): the sign symmetry in the forward mapping $x \mapsto x^2$ dictates that in the training set, there are frequent cases where x_i^2 and x_j^2 are close but x_i and x_j have different signs and are far apart. Although in theory neural networks with adequate capacities are universal function approximators, in practice they will struggle to learn such irregular functions. For general inverse problems, so long as the forward symmetries can relate remote inputs to the same output, such as all the three symmetries in PR, similar problems can surface.

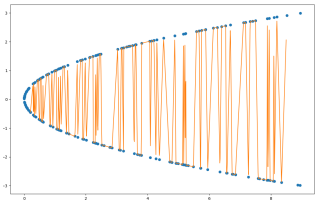

Figure 2: The highly oscillatory function defined by the training set when learning to take square root. Image credit: [17].

Biases in practical image datasets Strangely, the difficulty is only briefly touched in [16] among all recent work applying the end-to-end approach to PR. When we examine the training and test data that previous works use, it becomes clear that the issue is probably covered by intrinsic dataset biases. Previous experiments typically use images from standard computer vision datasets such as MNIST, ImageNet, CelebA(faces), where the image contents tend to be centralized and naturally oriented (see Fig. 3 (a)–(b)). This helps break the shift and flipping symmetries naturally, as these images are relatively close to each other compared to when mixed up with some of their symmetric copies. [1] Our analysis has been confirmed in a preliminary version of this work, where we showed that augmenting natural datasets to account for symmetries fails a state-of-the-art method that performs well without the augmentation [3]. In short, the fundamental difficulty has been concealed by biased data which do not reflect the essential properties of data in PR applications: nano-scale crystals (see Fig. 3 (c)), astronomical objects [20], where there is no natural orientation or centering of the image contents.

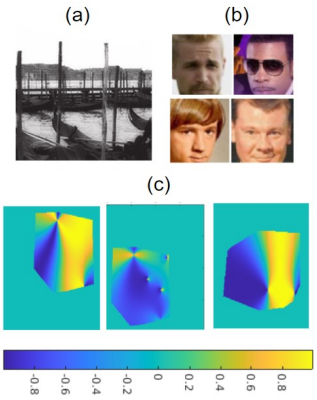

Figure 3: (a) & (b): Sample training images used in [13] and [15], respectively. (c) Sample images of simulated crystal structures in Bragg CDI applications [18,19] (only real parts of complex-valued images are shown), which do not have any natural orientation or centering.

2 Toward Practical PR

In this section, we sketch a two-stage method for solving PR, which consists of an initialization stage based on the end-to-end approach, and a refinement stage based on a robust iterative method.

2.1 Passive symmetry breaking

Let $\mathcal{T} \doteq \{\boldsymbol{X}_i, \boldsymbol{Y}_i = |\mathcal{F}(\boldsymbol{X})|^2\}$ be the training dataset we set up for implementing the end-to-end approach, and $g_{\boldsymbol{W}}$ be a chosen neural network parametrized by weights $\boldsymbol{W}$. Then the naive end-to-end approach will take the form $\min_{\boldsymbol{W}} \sum_i \ell(\boldsymbol{X}_i, g_{\boldsymbol{W}}(\boldsymbol{Y}_i))$, where ℓ is the loss function. Due to the symmetries, $\boldsymbol{Y}_i, \boldsymbol{Y}_j$ that are close may correspond to $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ that are centered at very different locations or flipped and hence far apart. This forces $g_{\boldsymbol{W}}$ to approximate a rapidly changing function, i.e., the difficulty that we alluded to above.

[1] The global phase symmetry is absent as they only deal with real-valued images.

The difficulty occurs because we require $g_{\boldsymbol{W}}(\boldsymbol{Y}_i)$ to match $\boldsymbol{X}_i$ as possible, where the latter induces mixed symmetries. How about we getting rid of $\boldsymbol{X}_i$'s? A natural alternative formulation is

$$\min_{\boldsymbol{W}} \sum_i \ell\Big(\boldsymbol{Y}_i, |\mathcal{F} \circ g_{\boldsymbol{W}}(\boldsymbol{Y}_i)|^2\Big), \tag{2.1}$$

as whether $g_{\boldsymbol{W}}(\boldsymbol{Y}_i)$ outputs $\boldsymbol{X}_i$ or any of its symmetric copies, $|\mathcal{F} \circ g_{\boldsymbol{W}}(\boldsymbol{Y}_i)|^2 \approx \boldsymbol{Y}_i \; \forall \; i$.

Why it might work? We toss away the difficult $\boldsymbol{X}_i$'s, but we also supply less information to the learning model. Now $g_{\boldsymbol{W}}$ has much more freedom, and it can still generate distant outputs for nearby inputs. Why is there hope? We draw our inspiration from the growing pile of evidence that neural networks optimized with first-order stochastic methods tend to learn simple functions over complicated ones, known as *implicit regularization* [21]. For our problem, $g_{\boldsymbol{W}}$ is simple when all the symmetries are broken and complicated when there are symmetries. So if implicit regularization occurs, symmetries are naturally broken.

Precursors Eq. (2.1) can be considered as an autoencoder objective with a known decoder. It is also similar to the cycle consistency idea [22–24] used in deep learning for several computer vision/graphics tasks. A unified theme is to approximate identity maps. But our motivation here is for implicit symmetry breaking taking advantage of implicit regularization, which differ from all other works. The same formulation and its equivalent form in the autocorrelation form has been independently proposed in [16]. They have not articulated the learning difficulty caused by all symmetries but the shift. Our previous work [17] proposes proprocessing steps for the training set to break symmetries for Gaussian PR—simplified version of PR where the shift and flipping symmetries are erased. The method can be generalized for PR as we show in a companion paper [25], but here we follow a different routine that seems simpler.

Jacobian regularization When the underlying distribution is not densely sampled, the learning difficulty is (counterintuitively) ameliorated, as the function determined by the data points becomes less oscillatory. This can be seen from Fig. 2 again and generalized to PR. Although complete symmetry breaking probably still corresponds to the "simplest" function, partial symmetry breaking leads to functions that are not worse off. This could lead to competitive local solutions to Eq. (2.1). To promote the simplest solution and hence complete symmetry breaking, we can optionally add in certain regularization terms to enhance regularity of the solution. Empirically, we find the Jacobian regularization [26] that encourages slowly varying functions helps, leading to the regularized objective

$$\min_{\boldsymbol{W}} \sum_i \ell\Big(\boldsymbol{Y}_i, |\mathcal{F} \circ g_{\boldsymbol{W}}(\boldsymbol{Y}_i)|^2\Big) + \lambda \|\boldsymbol{J}_g(\boldsymbol{Y}_i)\|_F^2. \tag{2.2}$$

To save computation, we adopt the approximation scheme proposed in [26] for the Jacobian term.

2.2 Refinement

The end-to-end learning performs prediction based on regression, the quality of which is often only mediocre and depends on how densely the underlying data distribution can be sampled. For many scientific inverse problems such as PR, high quality solution for each instance is always sought. In this work we propose to perform subsequent refinement using iterative methods. For PR, we observe our first stage produces results that often lead to reasonable support estimation, which suffices as initialization for popular iterative methods to solve PR. Here, we choose a recently proposed iterative method based on second-order augmented Lagrangian method, dubbed ALM [27], that exhibits fast convergence and good stability.

3 Experiments

Data We conduct our experiments on the FashionMNIST dataset [28], which is used by several previous works on PR. We take their $60,000$ training images and $10,000$ test images to construct our training and test sets, respectively. The images are 28×28. To simulate the typical black ground that causes the translation freedom in PR applications, we place all the images in a black background of 36×36. So $n = 36$, and we take $m = 72$ here to ensure injectivity of the forward model $2n - 1 =$

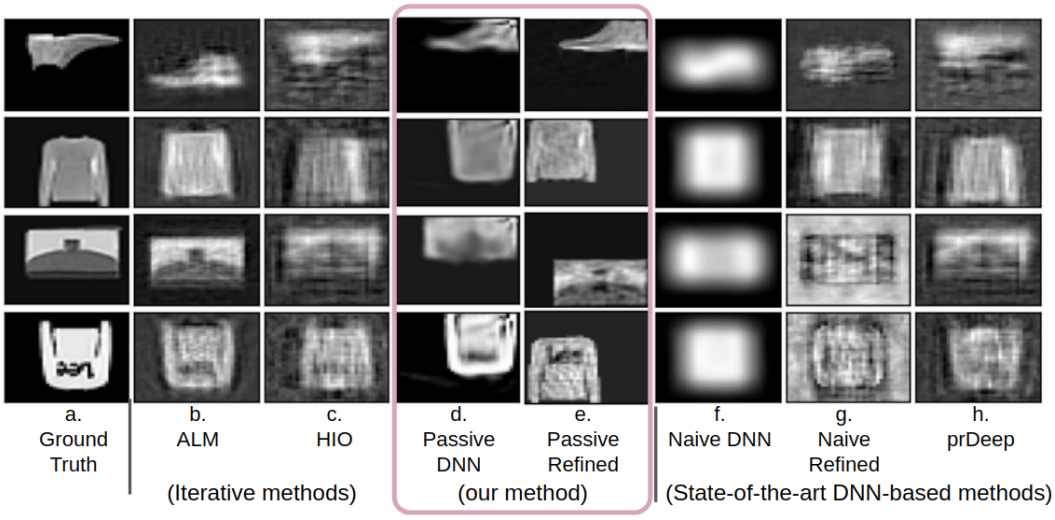

Figure 4: Visual comparison of our methods with several state-of-the-art methods for PR.

71 is exceeded. We further modify both train and test images with a random flipping and random translation operation to test the impact of symmetries on learning; samples are shown in Fig. 4-(a).

We use U-Net [29] (DNN) as our backbone DNN. We term our method (i.e., the two-stage method described in Section 2) **Passive Refined**. We compare it with HIO [5], ALM [27], Naive DNN [30], Passive DNN (DNN trained using in Eq. (2.2)), Naive Refined (Naive DNN refinement with ALM), prDeep [10]. HIO and ALM are representive iterative methods, and Naive DNN and prDeep are two of the state-of-the-art methods based on deep learning: Naive DNN implements the end-to-end approach, and prDeep is a hybrid approach that integrates deep learning and an iterative method, initialized with HIO.

Reconstruction results Results on randomly selected test images are presented in Fig. 4. Due to symmetries, both the iterative methods (ALM, HIO) and Naive DNN fail to recover the original images. The reconstruction from the passive DNN captures the object's shape, although the image details are not completely recovered. Thus, there is a need for additional refinement. We observe that the shape information from the first stage can be used to derive relatively accurate support of the object, which is ideal for initializing refinement stage which is based on classic iterative methods. After refinement, the final image is a reasonably accurate reconstruction of the original image.

Table 1 provides the average MSE adjusted to the symmetries (defined in Appendix A.1) for the test set. As noted, in the presence of symmetries, we see a huge gap in MSE of the reconstructed images between Passive Refined and other competing methods, consistent with the visual results.

The high-quality solution by our Passive Refined method depends crucially on good support estimation enabled by our Passive DNN stage. However, Naive DNN and the classic HIO—which is used to initialize prDeep [10], are unable to produce good support estimation to enable further refinement.[2]

Table 1: MSE error

	MSE
ALM	0.312
HIO	0.441
Passive DNN	0.266
Passive Refined	0.187
Naive DNN	0.492
Naive Refined	0.397
prDeep	0.412

Acknowledgments and Disclosure of Funding

KT and VK are supported by NSF BIGDATA 1838159. The authors acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the research results reported within this paper.

[2]The prDeep method assumes nonnegative real-valued images and makes use of the nonnegativity constraint in their algorithm. We have removed the nonnegativity constraint in their implementation for fair comparison, because our target is complex-valued images that are of central interest in imaging applications of PR.

References

[1] T. Bendory, R. Beinert, and Y. C. Eldar, "Fourier phase retrieval: Uniqueness and algorithms," in *Compressed Sensing and its Applications*. Springer International Publishing, 2017, pp. 55–91.

[2] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 87–109, may 2015.

[3] R. Manekar, K. Tayal, V. Kumar, and J. Sun, "End-to-end learning for phase retrieval," *ICML workshop on ML Interpretability for Scientific Discovery*, 2020.

[4] M. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 30, no. 2, pp. 140–154, apr 1982.

[5] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Applied Optics*, vol. 21, no. 15, p. 2758, aug 1982.

[6] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and fienup variants: a view from convex optimization," *Journal of the Optical Society of America A*, vol. 19, no. 7, p. 1334, jul 2002.

[7] V. Elser, I. Rankenburg, and P. Thibault, "Searching with iterated maps," *Proceedings of the National Academy of Sciences*, vol. 104, no. 2, pp. 418–423, jan 2007.

[8] D. R. Luke, S. Sabach, and M. Teboulle, "Optimization on spheres: Models and proximal algorithms with computational performance comparisons," *SIAM Journal on Mathematics of Data Science*, vol. 1, no. 3, pp. 408–445, jan 2019.

[9] M. Guizar-Sicairos and J. R. Fienup, "Understanding the twin-image problem in phase retrieval," *Journal of the Optical Society of America A*, vol. 29, no. 11, p. 2367, oct 2012.

[10] C. A. Metzler, P. Schniter, A. Veeraraghavan, and R. G. Baraniuk, "prdeep: Robust phase retrieval with a flexible deep network," *arXiv preprint arXiv:1803.00212*, 2018.

[11] Ç. Işıl, F. S. Oktem, and A. Koç, "Deep iterative reconstruction for phase retrieval," *Applied Optics*, vol. 58, no. 20, p. 5422, jul 2019.

[12] G. Ongie, A. Jalal, C. A. M. R. G. Baraniuk, A. G. Dimakis, and R. Willett, "Deep learning techniques for inverse problems in imaging," *IEEE Journal on Selected Areas in Information Theory*, 2020.

[13] A. Goy, K. Arthur, S. Li, and G. Barbastathis, "Low photon count phase retrieval using deep learning," *Physical Review Letters*, vol. 121, no. 24, dec 2018.

[14] M. J. Cherukara, Y. S. G. Nashed, and R. J. Harder, "Real-time coherent diffraction inversion using deep generative networks," *Scientific Reports*, vol. 8, no. 1, nov 2018.

[15] T. Uelwer, A. Oberstraß, and S. Harmeling, "Phase retrieval using conditional generative adversarial networks," *arXiv:1912.04981*, 2019.

[16] C. A. Metzler, F. Heide, P. Rangarajan, M. M. Balaji, A. Viswanath, A. Veeraraghavan, and R. G. Baraniuk, "Deep-inverse correlography: towards real-time high-resolution non-line-of-sight imaging," *Optica*, vol. 7, no. 1, p. 63, jan 2020.

[17] K. Tayal, C.-H. Lai, V. Kumar, and J. Sun, "Inverse problems, deep learning, and symmetry breaking," *arXiv preprint arXiv:2003.09077*, 2020.

[18] S. Maddali, P. Li, A. Pateras, D. Timbie, N. Delegan, A. L. Crook, H. Lee, I. Calvo-Almazan, D. Sheyfer, W. Cha, F. J. Heremans, D. D. Awschalom, V. Chamard, M. Allain, and S. O. Hruszkewycz, "General approaches for shear-correcting coordinate transformations in bragg coherent diffraction imaging. part i," *Journal of Applied Crystallography*, vol. 53, no. 2, pp. 393–403, mar 2020.

[19] P. Li, S. Maddali, A. Pateras, I. Calvo-Almazan, S. Hruszkewycz, W. Cha, V. Chamard, and M. Allain, "General approaches for shear-correcting coordinate transformations in bragg coherent diffraction imaging. part II," *Journal of Applied Crystallography*, vol. 53, no. 2, pp. 404–418, mar 2020.

[20] J. R. Fienup, "Phase retrieval for image reconstruction," in *Imaging and Applied Optics 2019 (COSI, IS, MATH, pcAOP)*. OSA, 2019.

[21] B. Neyshabur, R. Tomioka, and N. Srebro, "In search of the real inductive bias: On the role of implicit regularization in deep learning," *arXiv preprint arXiv:1412.6614*, 2014.

[22] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[23] C. Godard, O. Mac Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 270–279.

[24] T. Zhou, P. Krahenbuhl, M. Aubry, Q. Huang, and A. A. Efros, "Learning dense correspondence via 3d-guided cycle consistency," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 117–126.

[25] K. Tayal, C.-H. Lai, R. Manekar, Z. Zhuang, V. Kumar, and J. Sun., "Unlocking inverse problems using deep learning: Breaking symmetries in phase retrieval," in *NeurIPS 2020 Workshop on Deep Learning and Inverse Problems*, 2020.

[26] J. Hoffman, D. A. Roberts, and S. Yaida, "Robust learning with jacobian regularization," *arXiv preprint arXiv:1908.02729*, 2019.

[27] Z. Zhuang, G. Wang, Y. Travadi, and J. Sun, "Phase retrieval via second-order nonsmooth optimization," *ICML workshop on Beyond First Order Methods in Machine Learning*, 2020.

[28] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.

[29] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[30] A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," *Optica*, vol. 4, no. 9, p. 1117, sep 2017.

A Appendix

A.1 Mean Square Error (MSE)

Our reconstructed image is in $\mathbb{C}^{m\times m}$, where our original image is in $\mathbb{C}^{n\times n}$. To account for the three symmetries when taking MSE measure, we take the following steps: we take the original image, and scan through the larger reconstructed image to account for the translation symmetry. At each scan position, we calculate an adjusted MSE between the current patch $\boldsymbol{B} \in \mathbb{C}^{n\times n}$ and the original image $\boldsymbol{A}$. A $\lambda > 0$ and a global phase factor $e^{i\theta}$ (to account for the global phase) are introduced when calculating the MSE, i.e.,

$$\min_{\theta,\eta\geq 0} \left\|\boldsymbol{A} - \eta \boldsymbol{B} e^{i\theta}\right\|_F^2. \tag{A.1}$$

The smallest adjusted MSE is recorded over all scan positions. Then, the original image $\boldsymbol{A}$ is 2D flipped and conjugated and the same scanning process is repeated to calculate another smallest MSE, to account for the flipping symmetry. The smaller of the smallest MSE values is finally taken.

Below, we show that the optimal value in Eq. (A.1) can be easily computed. First we expand the square inside the objective and perform partial minimization with respect to θ, leading to

$$\max_{\theta} \operatorname{Re}\left\langle \boldsymbol{A}, \boldsymbol{B}e^{i\theta}\right\rangle. \tag{A.2}$$

But $\operatorname{Re}\left\langle \boldsymbol{A}, \boldsymbol{B}e^{i\theta}\right\rangle = \operatorname{Re}\left(\langle \boldsymbol{A}, \boldsymbol{B}\rangle\, e^{i\theta}\right) \leq \left|\langle \boldsymbol{A}, \boldsymbol{B}\rangle\, e^{i\theta}\right| \leq |\langle \boldsymbol{A}, \boldsymbol{B}\rangle|$ and the upper bound is achievable when $\theta = -\angle\langle \boldsymbol{A}, \boldsymbol{B}\rangle$. So the optimization problem now becomes

$$\min_{\eta\geq 0} \|\boldsymbol{A}\|_F^2 + \eta^2\|\boldsymbol{B}\|_F^2 - 2\eta|\langle \boldsymbol{A}, \boldsymbol{B}\rangle|. \tag{A.3}$$

The minimum of Eq. (A.3) occurs either when $\eta = 0$, which is $\|\boldsymbol{A}\|_F^2$, or by unconstrained optimality condition when $2\eta\|\boldsymbol{B}\|_F^2 = 2|\langle \boldsymbol{A}, \boldsymbol{B}\rangle| \implies \eta = |\langle \boldsymbol{A}, \boldsymbol{B}\rangle|/\|\boldsymbol{B}\|_F^2$, leading to the function value

$$\|\boldsymbol{A}\|_F^2 - \frac{|\langle \boldsymbol{A}, \boldsymbol{B}\rangle|}{\|\boldsymbol{B}\|_F^2}, \tag{A.4}$$

which is the smaller one.