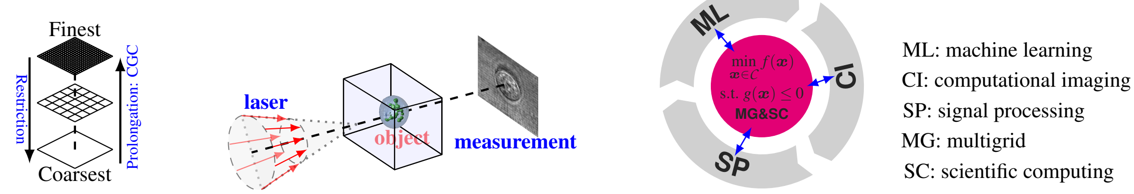
Research Statement

Tao Hong *hongtao.github.io*

Computational methods (CMs) naturally find applications in all fields of engineering and sciences. With the tremendous growth in computing power over the past decades, researchers can now build more accurate models to better understand the real world. These models often exhibit nonlinearity, nonconvexity, and huge-scale characteristics, requiring efficient CMs to save time and energy. My research **interests** lie in ***numerical*** optimization and ***multigrid*** computational methods. My research ***goal*** is to develop ***efficient*** CMs and software for scientific computing (SC), machine learning (ML), signal processing (SP), and computational imaging (CI).



(a) Multigrid (b) Optical diffraction tomography (c) Roadmap

Figure 1: (a-b) Principles of mutigrid and optical diffraction tomography; (c) My research roadmap.

Completed Research

Acceleration methods strive to boost computational efficiency, overcome obstacles in large-scale problems, real-time applications, and resource-limited settings, ultimately saving resources and enabling practical solutions to complex issues in diverse domains [13]. In my doctoral and postdoctoral research, I developed several impactful works on acceleration methods and studied efficient computational techniques for computational imaging (CI) [1–7]. Compressive sensing (CS) enables the recovery of signals from a small number of measurements, revolutionizing data acquisition and compression [14]. During my doctoral research, I made noteworthy contributions to optimizing CS systems, aiming to improve signal recovery accuracy and sampling efficiency by designing effective sampling strategies and sparse dictionaries [8–11].

Acceleration Methods

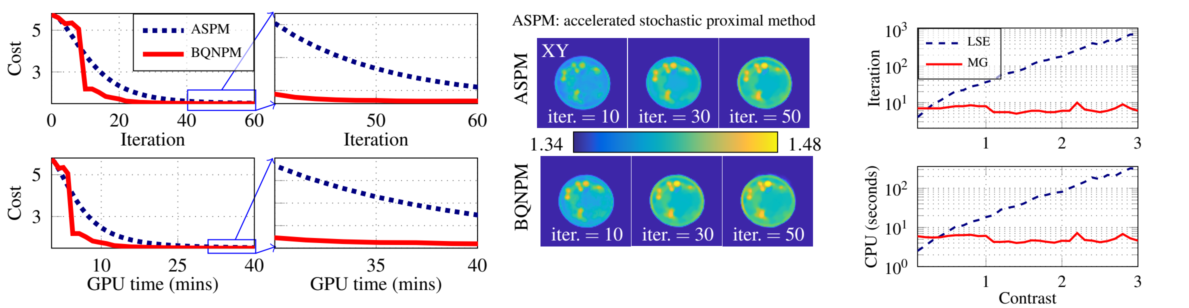
Plain (unaccelerated) iterative methods (PIMs) can be expedited by leveraging previous iterates with minimal computation [17, 18]. Nesterov's scheme (NS) is widely employed to accelerate various PIMs [19–21]. However, adapting NS for abstract PIMs remains an open question, and the development of a universal acceleration scheme for hierarchical problems is still understudied.

Adapting Nesterov's scheme [1]. Iterative solutions are often required in scientific computing for systems of the form $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{f}$, where $\boldsymbol{A}$ is a sparse, large-scale, and ill-conditioned matrix [15, 22]. Compared with classic Krylov subspace acceleration methods, NS (defined at the kth iteration as $\boldsymbol{x}_{k+1} = \boldsymbol{B}(\boldsymbol{u}_k)$, $\boldsymbol{u}_{k+1} = (1+c_k)\boldsymbol{x}_{k+1} - c_k\boldsymbol{x}_k$, where $\boldsymbol{B}$ denotes some PIM) requires more iterations but has the advantages of straightforward implementation, reduced memory requirements, and similar computational costs as PIMs. My work [1] adapted NS to accelerate PIMs for $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{f}$ and sought a closed-form solution for the optimal c_k. I also derived the asymptotic convergence factor (ACF) and studied the robustness of NS for certain $\boldsymbol{B}$ with complex eigenvalues.

Merging multigrid optimization with SESOP [2]. Multigrid methods (MGMs) [22, 23] are efficient solutions for solving elliptic partial differential equations (PDEs) and hierarchical problems, leveraging hierarchical grids and multiple scales. Fig. 1(a) presents the hierarchical grids of MGMs. However, the design of efficient MGMs presents a challenge, leading to the frequent combination of MGMs with acceleration techniques [24]. Combining acceleration with imperfect coarse grid correction (CGC) also poses challenges. SESOP is a framework that addresses large-scale unconstrained minimization problems by performing sequential optimization over affine subspaces ($\mathcal{M}_k$) formed by the current and past directions [25]. My work [2] integrated MGMs with SESOP (termed SESOP-MG) to accelerate MGMs while considering imperfect CGC. I also examined the ACF of SESOP-MG in its two-grid version and estimated the acceleration using the h-ellipticity measure [26]. For linear problems, I incorporated three directions (the preconditioned gradient, history, and CGC) in $\mathcal{M}_k$ and derived optimal fixed stepsizes for these directions.

Acceleration, Second-order Methods, and Multigrid in Imaging Sciences

Inverse problems (IPs) find diverse applications in engineering and sciences, enabling the estimation of parameters from indirect measurements. However, the majority of IPs in numerous imaging applications are highly *ill-posed*, *large-scale*, and *nonlinear*, resulting in complex optimization problems that require highly efficient and scalable numerical solvers.



(a) Cost versus iteration and GPU time (b) Reconstructed XY-plane images (c) Iterations and CPU time vs. contrast

Figure 2: (a-b) Performance on the 3D real data (yeast cell); (c) Performance on a disk sample.

Acceleration in regularization by denoising (RED) [3, 4]. RED [27] is an attractive framework for solving IPs by incorporating state-of-the-art denoisers as priors. A drawback of this approach is that the high computation of denoisers dominates the computation time. My work [3] used vector extrapolation methods [28] to accelerate the algorithms presented in [27]. That was done by solving an additional very small-scale quadratic minimization problem every few iterations with almost no cost, reducing the frequency of executing denoisers. My subsequent work [4] explored the utilization of weighted proximal methods (WPMs) in RED, incorporating second-order information and achieving an approximate 10-fold speedup compared with the best algorithm proposed in [27].

Second-order methods in CS magnetic resonance imaging (MRI) and optical diffraction tomography (ODT) [5, 6]. IPs are often modeled as large-scale composite minimization problems ($\min_{\boldsymbol{x}} f(\boldsymbol{x}) + g(\boldsymbol{x})$), favoring the use of first-order methods (FOMs) [29]. Second-order methods (SOMs) are faster than FOMs in terms of iterations but receive limited attention in IPs due to the need for solving $\operatorname{prox}_g^{\boldsymbol{W}}(\boldsymbol{u}) = \arg\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{u}\|_{\boldsymbol{W}}^2 + g(\boldsymbol{x}), \boldsymbol{W} \succ 0$ at each iteration, unlike FOMs where $\operatorname{prox}_g^{\boldsymbol{I}}(\boldsymbol{u})$ often has a closed-form solution. CS MRI allows the reconstruction of *complex* MRI images ($\boldsymbol{x} \in \mathbb{C}^N$) from undersampled data, offering potential for accelerated imaging [30]. My work [5] introduced a complex quasi-Newton proximal method (CQNPM) for CS MRI reconstruction when wavelet and total variation regularizers are both used and proposed efficient approaches to solve $\operatorname{prox}_g^{\boldsymbol{W}}(\boldsymbol{u})$. CQNPM outperforms FOMs in terms of iterations and CPU time.

Three-dimensional ODT (Fig. 1(b)) [31] is a noninvasive imaging modality for recovering the refractive index of an object from indirect measurements. ODT reconstruction involves minimizing

$\min_{\boldsymbol{x}\in\mathcal{C}}\sum_{m=1}^{M} f_m(\boldsymbol{x})+g(\boldsymbol{x})$, $M \gg 1$, which is challenging due to the nonlinearity and nonconvexity of $\{f_m\}_{m\geq 1}$, the high computation of $\{\nabla f_m\}_{m\geq 1}$, the large size of M, the nonsmoothness of $g(\boldsymbol{x})$, and the constraint. My work [6] proposed a mini-batch quasi-Newton proximal method (BQNPM) such that the computation at each iteration is independent of M and developed an efficient approach to solve the related $\mathrm{prox}_g^{\boldsymbol{W}}(\boldsymbol{u}) = \arg\min_{\boldsymbol{x}\in\mathcal{C}} \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{u}\|_{\boldsymbol{W}}^2 + g(\boldsymbol{x})$. BQNPM outperforms stochastic FOMs in terms of iterations and GPU time, see Fig. 2(a) and Fig. 2(b).

Multigrid in diffraction tomography (DT) [7]. CI with accurate nonlinear physical models attracts interest for high-quality reconstructions. The Lippmann-Schwinger equation (LSE) forward model [31] enables high-contrast object recovery in DT. However, the numerical solvers for LSE with high-contrast object are inefficient, hindering efficient reconstruction. My work [7] suggested using the Helmholtz equation as the forward model, which is as accurate as LSE. We also introduced an efficient and scalable geometric ***multigrid*** solver for the Helmholtz equation, accelerating high-contrast object recovery by efficiently solving the Helmholtz equation, see Fig. 2(c).

Optimizing Compressive Sensing Systems

Compressive sensing (CS) allows for the recovery of a sparse signal $\boldsymbol{x}\in\mathbb{R}^N := \boldsymbol{D}\boldsymbol{\theta}+\boldsymbol{e}, \boldsymbol{D}\in\mathbb{R}^{N\times L}$, $\|\boldsymbol{\theta}\|_0 \ll L$ from incomplete measurements $\boldsymbol{y}\in\mathbb{R}^M := \boldsymbol{\Phi}\boldsymbol{x}$ ($M \ll N$) [14]. Under the assumption that $\boldsymbol{e}=\boldsymbol{0}$, Elad [32] showed that refining $\boldsymbol{\Phi}$ can enhance recovery accuracy. However, practical signals with $\boldsymbol{e}\neq\boldsymbol{0}$ pose challenges for optimizing $\boldsymbol{\Phi}$. In my work [8], I proposed a new framework to refine $\boldsymbol{\Phi}$ by jointly minimizing the coherence of $\boldsymbol{\Phi}\boldsymbol{D}$ and $\|\boldsymbol{\Phi}\boldsymbol{e}\|_2^2$, which is robust to $\boldsymbol{e}\neq\boldsymbol{0}$. Furthermore, my subsequent study [8] improved the efficiency of refining $\boldsymbol{\Phi}$ by eliminating the requirement of prior knowledge $\boldsymbol{e}$. In later research [10], I suggested refining $\boldsymbol{\Phi}$ and $\boldsymbol{D}$ simultaneously for high-dimensional signals by using an online algorithm. Lastly, in my work [11], I considered the sampling efficiency by optimizing a structured $\boldsymbol{\Phi}=\bar{\boldsymbol{\Phi}}\boldsymbol{\Psi}$ where $\bar{\boldsymbol{\Phi}}$ is extremely sparse and $\boldsymbol{\Psi}$ possesses fast transform properties (e.g., wavelets) that require fewer arithmetic operations to sample signals. My contributions in this area offer new perspectives on optimizing MRI sampling trajectories, opening up avenues for future research.

Future Research Plans

Accurate models and efficient algorithms are essential for managing the vast amount of data generated in engineering and sciences. My future work will concentrate on designing accurate and efficient models and algorithms to tackle challenges in areas such as SC, SP, ML, and computational imaging (CI). My complete research roadmap is summarized in Fig. 1(c).

Optimization with multigrid and computational imaging. Optimization is fundamental to maximizing efficiency and enhancing decision-making in various fields. Compared with first-order methods, high-order methods offer accelerated convergence but face various practical issues, such as increased computational complexity, global convergence, and limitation to nonsmooth problems. I plan to study the application of high-order methods to various scenarios, delve into nonsmooth and nonconvex optimization, and explore their theoretical underpinning.

- *High-order nonlinear multigrid.* The full approximation storage (FAS) scheme, widely used in nonlinear multigrid methods (MGMs), leverages only first-order information [22]. While the benefits of using second-order information have been introduced in [33], this approach has not been broadly applied due to the technical challenges in making it robust and efficient. My plan is to explore ways to use high-order information in nonlinear MGMs under convex optimization frameworks [34, 35] and to study the convergence of high-order nonlinear MGMs. The studies in this direction can potentially revolutionize the solution of nonlinear PDEs using MGMs.

- *Nonsmooth and nonconvex optimization for computational imaging.* Computational imaging (CI) integrates imaging techniques and computational methods to enhance image acquisition, processing, and analysis, thereby improving image quality and advancing capabilities in various domains such as microscopy imaging [36], medical imaging [37], and geophysics [38] etc. The core computational task in CI is solving (nonlinear) inverse problems, many of which are nonconvex (e.g., ODT [6]), nonsmooth (e.g., intensity ODT [38]), and computationally expensive (e.g., large-scale 3D images [31]). My goal is to deepen our understanding of these inverse problems by establishing recovery guarantees, thereby enabling the development of new imaging systems and efficient optimization algorithms, as exemplified in my work [4–6]. I am enthusiastic about collaborating across domains, utilizing nonsmooth and nonconvex optimization to solve large-scale problems, and contributing to open software packages, as demonstrated in my previous research. Moreover, this line of study holds potential to contribute to the theoretical understanding of training deep neural networks, a complex nonsmooth and nonconvex problem.
- *Adapting Nesterov's scheme to nonlinear problems.* We have seen the merits of using Nesterov's scheme (NS) to accelerate PIMs for linear problems [1]. However, adapting NS to nonlinear problems remains an unexplored area. I plan to study efficient and practical ways to adapt NS to accelerate abstract solvers for nonlinear problems, which will provide new insights for acceleration methods.

Efficient deep neural networks (DNNs). DNNs have been prominent in AI since 2012 [42]. Given the large size of modern DNNs, there is a necessity for efficient DNNs to achieve faster, cost-effective, and eco-friendly AI. I plan to focus on designing small and efficient physics-informed NNs for diverse applications by using domain knowledge. Quantizing parameters with fewer bits can compress NNs and save computation [43]. Matrix-vector multiplication (MCM), a fundamental building block in DNNs, can have infinite realizations that impact quantization error (QE) [44]. My second aim is to study MCM realizations to minimize the effects of QE, thereby reducing bit usage while preserving performance, as in my prior work on all-pass digital filter structures [12]. Furthermore, I intend to develop theoretical frameworks to guide the discovery of realizations for different NNs. I am also eager to collaborate with hardware specialists to apply our findings to real edge devices, smartphones, and embedded systems.

ASL from theory to practice. Arterial spin labeling (ASL) [39] is a non-invasive MRI technique that measures cerebral blood flow by magnetically labeling arterial blood, providing valuable information about brain perfusion. However, ASL techniques face challenges such as low SNR, and are susceptible to confounding factors like pulsatility, motion, and variations in bolus arrival time (BAT). As a continuation of my postdoctoral research, I plan to study the combination of velocity-selective inversion pulses with MR fingerprinting (MRF) [40]. This approach allows us to estimate multiple hemodynamic parameters (MHPs) in a single scan and reduces the signal loss caused by BAT. Moreover, I aim to globally optimize the ASL-MRF scan, acquisition, reconstruction, and parameter estimation processes, which can shorten scan time while still accurately obtaining MHPs. However, the whole optimization is nonlinear, nonconvex, and nonsmooth, which requires efficient and effective computational methods. I also intend to collaborate with neuroscientists and obstetricians to apply ASL techniques to the study of Parkinson's disease and placental imaging. I believe that the non-invasive, quantitative, repeatable measurements, and cost-effective properties of ASL will make it a popular technique for early diagnosis of diseases in the brain and fetus.

References

Part I – My Selected Publications (* equal contribution)

[1] **Tao Hong** and Irad Yavneh. On adapting Nesterov's scheme to accelerate iterative methods for linear problems. *Numerical Linear Algebra with Applications*, 29(2):e2417, 2022. [link] [pdf] [code][slides].

[2] **Tao Hong**, Irad Yavneh, and Michael Zibulevsky. Merging multigrid optimization with SESOP, *In Preparation for Submitting to SIAM Journal on Scientific Computing*, 2023. [link][pdf][slides].

[3] **Tao Hong**, Yaniv Romano, and Michael Elad. Acceleration of RED via vector extrapolation. *Journal of Visual Communication and Image Representation*, 63:102575, 2019. [link][pdf][code].

[4] **Tao Hong**, Irad Yavneh, and Michael Zibulevsky. Solving RED with weighted proximal methods. *IEEE Signal Processing Letters*, 27:501-505, 2020. [link][pdf][code][slides].

[5] **Tao Hong**, Luis Hernandez-Garcia, and Jeffrey A Fessler. A complex quasi-Newton proximal method for image reconstruction in compressed sensing MRI. *Submitted to IEEE Transactions on Computational Imaging,* 2023. [link][pdf][code][slides].

[6] **Tao Hong***, Thanh-an Pham*, Irad Yavneh, and Michael Unser. A mini-batch quasi-Newton proximal method for constrained total-variation nonlinear image reconstruction. *, Submitted to IEEE Transactions on Computational Imaging,* 2023. [link][pdf][poster].

[7] **Tao Hong***, Thanh-an Pham*, Eran Treister, and Michael Unser. Diffraction tomography with Helmholtz equation: Efficient and robust multigrid- based solver. *In Preparation for Submitting to Inverse Problems,* 2023. [link][pdf].

[8] **Tao Hong**, Huang Bai, Sheng Li, and Zhihui Zhu. An efficient algorithm for designing projection matrix in compressive sensing based on alternating optimization. *Signal Processing,* 125:9-20, 2016. [link][pdf][code].

[9] **Tao Hong** and Zhihui Zhu. An efficient method for robust projection matrix design. *Signal Processing,* 143:200-210, 2018. [link][pdf][code].

[10] **Tao Hong** and Zhihui Zhu. Online learning sensing matrix and sparsifying dictionary simultaneously for compressive sensing. *Signal Processing,* 153:188-196, 2018. [link][pdf][code].

[11] **Tao Hong**, Xiao Li, Zhihui Zhu, and Qiuwei Li. Optimized structured sparse sensing matrices for compressive sensing. *Signal Processing,* 159:119-129, 2019. [link][pdf].

[12] **Tao Hong**, Chaogeng Huang, Gang Li, and Yong Ching Lim. A Hessenberg-based input balanced realization for all-pass systems. *2013 9th International Conference on Information, Communications & Signal Processing.* IEEE, 2013. [pdf]

Part II – Other Publications

[13] Alexandre d'Aspremont, Damien Scieur, and Adrien Taylor, Acceleration methods. *Foundations and Trends® in Optimization*, 5(1-2):1-245, 2021.

[14] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine,* 25(2):21-30, 2008.

[15] Yousef Saad. *Iterative Methods for Sparse Linear Systems.* SIAM, 2003.

[16] Yurii Nesterov. *Lectures on Convex Optimization,* volume 137. Springer, 2018.

[17] Yurii Nesterov. A method for solving the convex programming problem with convergence rate $O(1/k^2)$. In *Dokl. akad. nauk Sssr,* 269:543-547, 1983.

[18] Youcef Saad and Martin H Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing,* 7(3):856-869, 1986.

[19] Yurii Nesterov. Accelerating the cubic regularization of Newton's method on convex problems. *Mathematical Programming,* 112(1):159-181, 2008.

[20] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences,* 2(1):183-202, 2009.

[21] Yurii Nesterov and Sebastian U Stich. Efficiency of the accelerated coordinate descent method on structured optimization problems. *SIAM Journal on Optimization*, 27(1):110-123, 2017.

[22] Achi Brandt. Multi-level adaptive solutions to boundary-value problems. *Mathematics of Computation*, 31(138):333-390, 1977.

[23] Irad Yavneh. Why multigrid methods are so efficient. *Computing in Science & Engineering*, 8(6):12-22, 2006.

[24] Cornelis W Oosterlee and Takumi Washio. Krylov subspace acceleration of nonlinear multigrid with application to recirculating flows. *SIAM Journal on Scientific Computing,* 21(5):1670-1690, 2000.

[25] Michael Zibulevsky. Speeding-up convergence via sequential subspace optimization: Current state and future directions. *arXiv preprint arXiv:1401.0159*, 2013.

[26] Achi Brandt and Oren E Livne. Multigrid Techniques: 1984 Guide with Applications to Fluid Dynamics, volume 67. SIAM, 2011.

[27] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (RED). *SIAM Journal on Imaging Sciences,* 10(4):1807-1844, 2017.

[28] Avram Sidi. *Vector Extrapolation Methods with Applications.* SIAM, 2017.

[29] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127-239, 2014.

[30] Michael Lustig, David L Donoho, Juan M Santos, and John M Pauly. Compressed sensing MRI, *IEEE Signal Processing Magazine,* 25(2):72-82, 2008.

[31] Thanh-an Pham, Emmanuel Soubies, Ahmed Ayoub, Joowon Lim, Demetri Psaltis, and Michael Unser. Three-dimensional optical diffraction tomography with Lippmann-Schwinger model. *IEEE Transactions on Computational Imaging,* 6:727-738, 2020.

[32] Michael Elad. Optimized projections for compressed sensing. *IEEE Transactions on Signal Processing,* 55(12):5695-5702, 2007.

[33] Irad Yavneh and Gregory Dardyk. A multilevel nonlinear method. *SIAM Journal on Scientific Computing,* 28(1):24-46, 2006.

[34] Yurii Nesterov and Boris T Polyak. Cubic regularization of Newton method and its global performance. *Mathematical Programming,* 108(1):177-205, 2006.

[35] Yurii Nesterov. Implementable tensor methods in unconstrained convex optimization. *Mathematical Programming,* 186:157-183, 2021.

[36] Mickaël Lelek, Melina T Gyparaki, Gerti Beliu, and Florian Schueder et al. Single-molecule localization microscopy. *Nature Reviews Methods Primers,* 1(1):39, 2021.

[37] Paul Suetens. *Fundamentals of Medical Imaging.* Cambridge University Press, 2017.

[38] Ludovic Métivier, Romain Brossier, Jean Virieux, and Stéphane Operto, Full waveform inversion and the truncated Newton method. *SIAM Journal on Scientific Computing,* 35(2):B401-B437, 2013.

[39] David C Alsop, John A Detre, and Xavier Golay et al. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. *Magnetic Resonance in Medicine,* 73(1):102-116, 2015.

[40] Dan Ma, Vikas Gulani, and Nicole Seiberlich et al. Magnetic resonance fingerprinting. *Nature*, 495(7440):187-192, 2013.

[41] Stephen P Jordan, Siyuan Hu, and Ignacio Rozada et al. Automated design of pulse sequences for magnetic resonance fingerprinting using physics-inspired optimization. *Proceedings of the National Academy of Sciences,* 118(40):e2020516118, 2021.

[42] Alex, Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems,* 25, 2012.

[43] Amir Gholami, Sehoon Kim, and Zhen Dong et al. A survey of quantization methods for efficient neural network inference. *In Low-Power Computer Vision,* 291-326, 2022.

[44] Michel Gevers and Gang Li. *Parametrizations in Control, Estimation and Filtering Problems: Accuracy Aspects*. Springer Science & Business Media, 1993.