Massimo Fornasier, Hui Huang, Lorenzo Pareschi, Philippe Sünnen

TU München

Fonds National de la Recherche Luxembourg

DFG Deutsche Forschungsgemeinschaft

Consensus-based Optimization on Hypersurfaces

Introduction

Goal: Global optimization of nonconvex functions on smooth and comp. hypersurfaces.

Setup: Let $\mathcal{E} : \Gamma \to \mathbb{R}_+$ where $\Gamma = \{x \in \mathbb{R}^d \mid \gamma(x) = 0\}$ and $v_\star = \arg\min_\Gamma \mathcal{E}$.

Model [4]: System of interacting particles

$$dV_t^i = \lambda P_\Gamma(V_t^i) V_t^\alpha dt + \sigma P_\Gamma(V_t^i) D_t^i dB_t^i - \frac{\sigma^2}{2}(|V_t^i - V_t^\alpha|^2 + (D_t^i)^2 - 2|D_t^i V_t^i|^2) V_t^i dt$$

- $D_t^i = \mathbf{diag}(V_t^i - V_t^\alpha)$ and the consensus point is $V_t^\alpha = \sum e^{-\alpha\mathcal{E}(V_t^i)} V_t^i / \sum e^{-\alpha\mathcal{E}(V_t^i)}$
- Projection $P_\Gamma(v) = I - \nabla\gamma(v) \otimes \nabla\gamma(v)$, e.g., for $\Gamma = \mathbb{S}^{d-1}$, $P_{\mathbb{S}^{d-1}}(v) = I - \frac{v \otimes v}{|v|^2}$

Method:

- Euler-Maruyama discretization

$$\begin{cases} \tilde{V}_{n+1}^i = V_n^i + \lambda P_\Gamma(V_n^i) V_n^\alpha \Delta t \sigma P_\Gamma(V_n^i) D_n^i \Delta B_n^i - \Delta t \frac{\sigma^2}{2}(|V_n^i - V_n^\alpha|^2 + (D_n^i)^2 - 2|D_n^i V_n^i|^2) V_n^i \\ V_{n+1}^i = \Pi_\Gamma(\tilde{V}_{n+1}^i) \end{cases}$$

- Projection $\Pi_\Gamma(v) = \arg\min_{\tilde{v}\in\Gamma} |v - \tilde{v}|$, e.g., for $\Gamma = \mathbb{S}^{d-1}$, $\Pi_{\mathbb{S}^{d-1}}(v) = v/|v|$

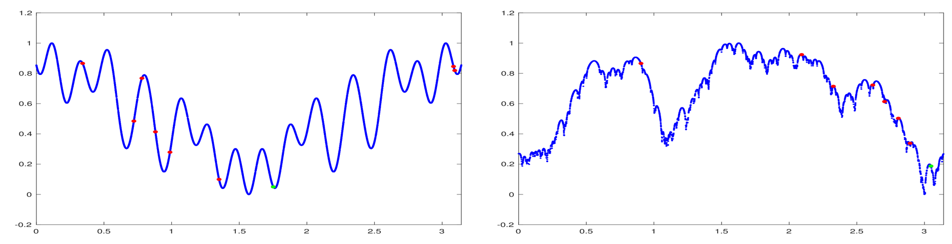

Fig. 1. Left: Rastrigin function on $\Gamma = \mathbb{S}^1$ with initial particles V_0^i (red) and consensus point V_t^α (green). Right: function $\mathcal{E}_p$ from the robust PCA application with $p = 0.1$.

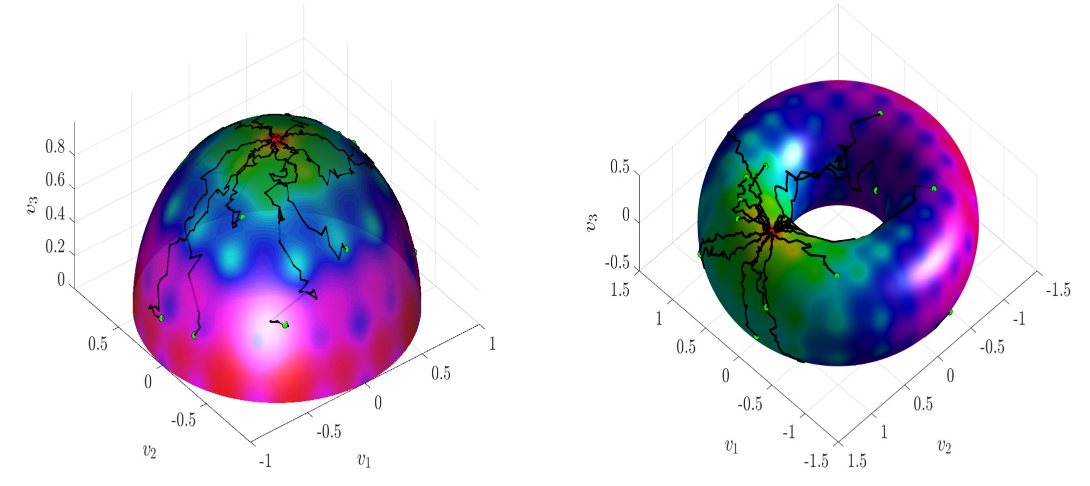

Fig. 2. Ackley function on $\mathbb{S}^2$ and on the torus $\mathbb{T}^2$.

Analysis

Sketch of the proof of convergence:

- Large particle limit $N \to \infty$: the system of first order SDE's approximates a deterministic PDE of mean-field type
- The solution $\rho_t(x)$ of the mean-field PDE converges to a delta function $\delta_{\bar{v}}$ as $T \to \infty$
- The delta function $\bar{v}$ is close to $v_\star$ if the initial data is *well-prepared*

Error Estimate:

- There is a set of parameters such that $|\mathbb{E}(\rho_{T^\star}) - v_\star| \leq \epsilon$
- We have $\mathbb{E}(|\frac{1}{N}\sum V^i_{\Delta t, n_{T^\star}} - v_\star|^2) \lesssim (\Delta t)^{2m} + N^{-1} + \epsilon^2$

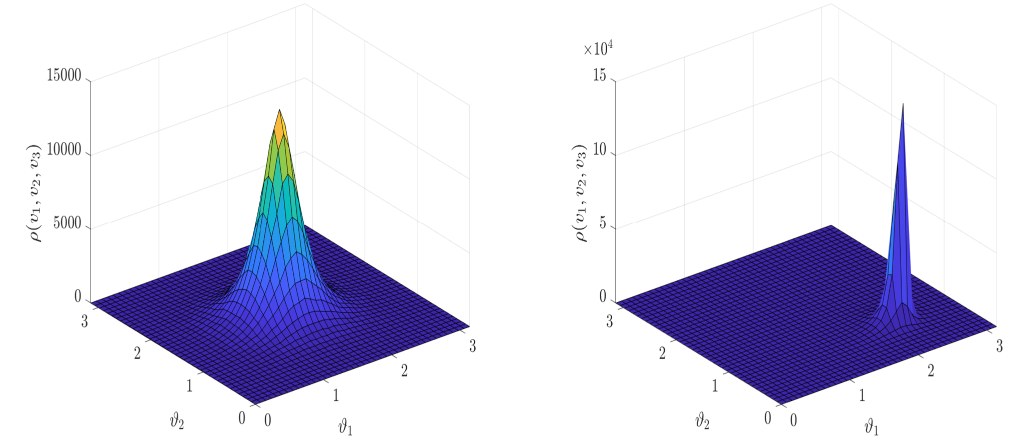

Fig. 3. The function $\rho_t(x)$ for different times t

Robust PCA

Setup: Point cloud $\mathbf{X} = \{x^{(i)} \in \mathbb{R}^d\}$ with cellwise and casewise contamination

Goal: Find principal component of $\mathbf{X}$ without weighting the outliers to much

Idea: Minimize $\mathcal{E}_p(v) = \sum_i |(I - v \otimes v)x^{(i)}|^p$ for $0 < p \leq 2$ with KV-CBO

Results:

- High accuracy for artificial data generated by Haystack model [3] in dimension $d = 100$
- Real data: robust computation of eigenfaces in dimension $d \approx 3000$

Fig. 4. Eigenfaces for *Extended Yale Face Database B* [7] and *10K US Adult Faces Database* [8]

Phase Retrieval

Setup: Measurements $y_i = |\langle z_\star, a^{(i)}\rangle|^2$ where $\{a^{(i)}\}_{i=1,\dots,M}$ is, e.g., a Gaussian frame

Goal: Reconstruct unknown vector $z_\star \in \mathbb{R}^d$ from known measurements $y \in \mathbb{R}^M$

Idea:

- Reformulation of the problem: $\tilde{y}_i = |\langle v_\star, \tilde{a}^{(i)}\rangle|^2$ with unknown $v_\star \in \mathbb{S}^d$
- Find $v_\star$ by minimizing the empirical risk $\mathcal{E}(v) = \sum_{i=1}^M |\langle \tilde{a}_i, v\rangle^2 - \tilde{y}_i|^2$ with KV-CBO
- Reconstruct $z_\star$ from $v_\star$

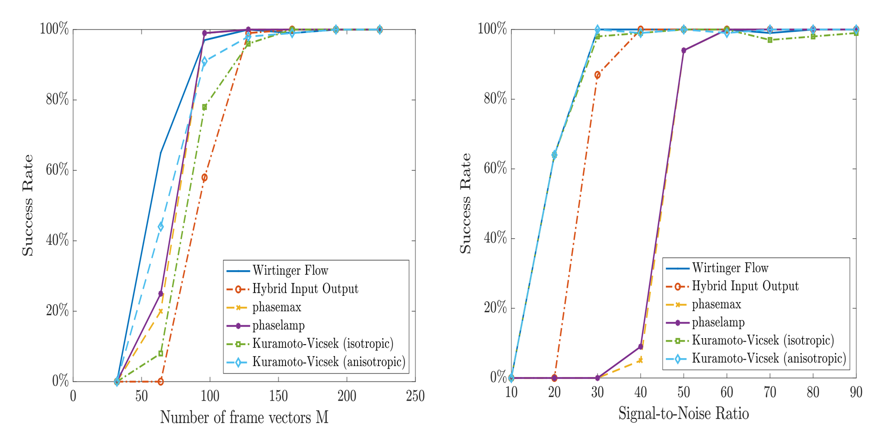

Fig. 5. Success rate in terms of number of frame vectors and signal-to-noise ratio for different benchmark methods.

Conclusion

- KV-CBO method is a consensus-based 0 order optimization method
- Convergence proof based on mean-field PDE
- No curse of dimensionality for $\Gamma = \mathbb{S}^{d-1}$ (conj. for any Γ)
- Computationally tractable.
- MATLAB implementation: `github.com/PhilippeSu/`

References

[1] R. Pinnau, C. Totzeck, O. Tse, and S. Martin. A consensus-based model for global optimization and its mean-field limit. Mathematical Models and Methods in Applied Sciences. 2017.

[2] J. A. Carrillo, Y.-Pil Choi, C. Totzeck, and O. Tse. An analytical framework for consensus-based global optimization method. Mathematical Models and Methods in Applied Sciences. 2018

[3] G. Lerman, M. McCoy, J. Tropp, T. Zhang. Robust Computation of Linear Models by Convex Relaxation. Foundations of Computational Mathematics. 2015.

[4] M. Fornasier, H. Huang, L. Pareschi, P. Sünnen. Consensus-based optimization on hypersurfaces: well-posedness and mean-field limit. Mathematical Models and Methods in Applied Sciences. 2020.

[5] M. Fornasier, H. Huang, L. Pareschi, P. Sünnen. Consensus based optimization on the sphere: Convergence to global mininizers and machine learning. arXiv:2001.11988. 2020.

[6] M. Fornasier, H. Huang, L. Pareschi, P. Sünnen. Anisotropic Diffusion in Consensus-based Optimization on the Sphere. in preparation.

[7] Lee, K.C., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans. Pattern Anal. Mach. Intell. 27(5), 684–698 (2005)

[8] Wilma. A. Bainbridge, Philipp Isola, and Aude Oliva. The Intrinsic Memorability of Face Photographs. Journal of Experimental Psychology: General, 142(4), 1323-1334., 2013.