Approximate Vanishing Ideal Computations at Scale

Elias Wirth [1] Hiroshi Kera [2] Sebastian Pokutta [1,3]

[1]Berlin Institute of Technology, [2]Chiba University, [3]Zuse Institute Berlin

Vanishing ideal

Given data set $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\} \subseteq \mathbb{R}^n$,

$$\mathcal{I}_X = \{f \in \mathbb{R}[x_1, \ldots, x_n] \mid f(\mathbf{x}) = 0 \ \forall \mathbf{x} \in X\},$$

the *vanishing ideal*, succinctly characterizes X. By Hilbert's basis theorem [1], there exists a finite number of *generators* $g_1, \ldots, g_k \in \mathcal{I}_X$, with $k \in \mathbb{N}$, such that for any $f \in \mathcal{I}_X$, there exist $h_1, \ldots, h_k \in \mathbb{R}[x_1, \ldots, x_n]$ such that

$$f = \sum_{i=1}^{k} g_i h_i.$$

Feature transformations with generators

Setting:

- Input space $\mathcal{X} \subseteq [-1, 1]^n$
- Output space $\mathcal{Y} = [k]$
- Training sample $S = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\} \in (\mathcal{X} \times \mathcal{Y})^m$ drawn *i.i.d.* from some unknown distribution $\mathcal{D}$

Goal:

- Determine a *hypothesis* $h \colon \mathcal{X} \to \mathcal{Y}$ with small *generalization error* $\mathbb{P}_{(\mathbf{x},y)\sim\mathcal{D}}[h(\mathbf{x}) \neq y]$

Pipeline:

- Let $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$
- For all $i \in [k]$, let $X^i \subseteq X$ denote the set of feature vectors corresponding to class i
- For all $i \in [k]$, construct a set of generators $\mathcal{G}^i = \{g_j^{(i)}\}_{j=1}^{|\mathcal{G}^i|}$ for the vanishing ideal $\mathcal{I}_{X^i}$
- Transform samples $\mathbf{x} \in X$ via the feature transformation
$$\mathbf{x} \mapsto \tilde{\mathbf{x}} = \left(\ldots, |g_1^{(i)}(\mathbf{x})|, \ldots, |g_{|\mathcal{G}^i|}^{(i)}(\mathbf{x})|, \ldots\right)^\intercal$$
- $\tilde{S} = \{(\tilde{\mathbf{x}}, y) \mid (\mathbf{x}, y) \in S\}$ is linearly separable
- Train a linear kernel SVM on $\tilde{S}$

Open question:

- How to construct the sets of generators $\mathcal{G}^i$?

Oracle approximate vanishing ideal algorithm

Algorithm 1 Oracle approximate vanishing ideal algorithm (OAVI)

Input: $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\} \subseteq \mathbb{R}^n$, $\psi \geq 0$, $\tau \geq 2$.
Output: $\mathcal{G}, \mathcal{O} \subseteq \mathbb{R}[x_1, \ldots, x_n]$.

$d \leftarrow 1$
$\mathcal{O} = \{t_1\}_\sigma \leftarrow \{\mathbb{1}\}_\sigma$
$\mathcal{G} \leftarrow \emptyset$
while $\partial_d \mathcal{O} = \{u_1, \ldots, u_k\}_\sigma \neq \emptyset$ **do**
  for $i = 1, \ldots, k$ **do**
    $P \leftarrow \{\mathbf{y} \in \mathbb{R}^{|\mathcal{O}|} \mid \|\mathbf{y}\|_1 \leq \tau - 1\}$
    $\mathbf{c} \in \operatorname{argmin}_{\mathbf{y} \in P} \frac{1}{m} \|\mathcal{O}(X)\mathbf{y} + u_i(X)\|_2^2$
    $g \leftarrow \sum_{j=1}^{|\mathcal{O}|} c_j t_j + u_i$
    if $\operatorname{mse}(g, X) \leq \psi$ **then**
      $\mathcal{G} \leftarrow \mathcal{G} \cup \{g\}$
    else
      $\mathcal{O} \leftarrow (\mathcal{O} \cup \{u_i\})_\sigma$
  $d \leftarrow d + 1$

- Generators vanishing on out-sample data [9]
- OAVI + linear kernel SVM inherit the margin bound of the SVM [9]
- Sparse generators

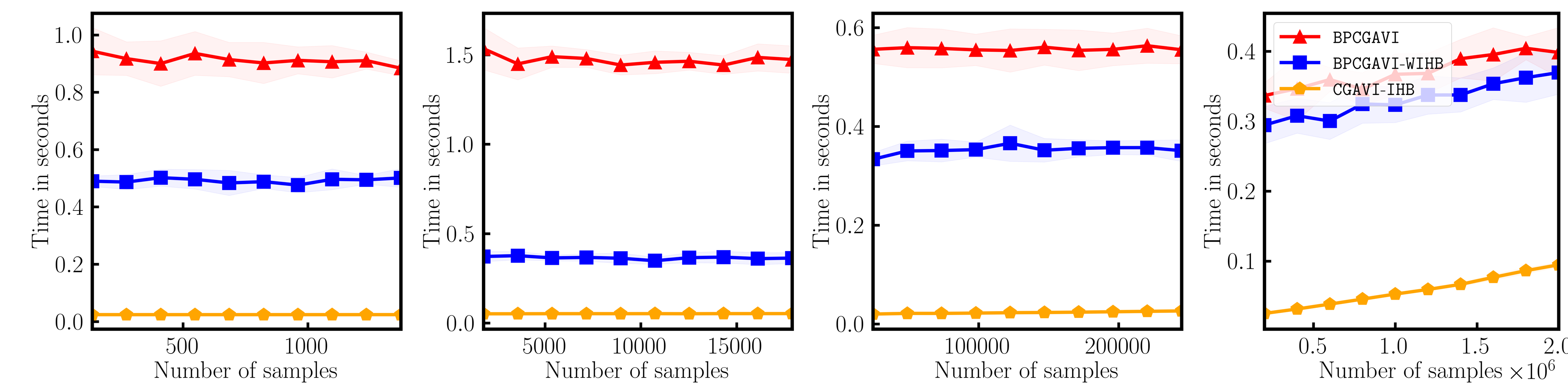
Running time of OAVI is linear in the number of samples m

Theorem. (Number-of-samples-agnostic bound on $|\mathcal{G}| + |\mathcal{O}|$.) Let $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\} \subseteq [0, 1]^n$, $\psi \in]0, 1[$, $D = \lceil -\log(\psi)/\log(4) \rceil$, $\tau \geq (3/2)^D$, and $(\mathcal{G}, \mathcal{O}) = \texttt{OAVI}(X, \psi, \tau)$. Then, OAVI terminates after having constructed generators of degree D. Thus, $|\mathcal{G}| + |\mathcal{O}| \leq \binom{D+n}{D}$.

Main result

For the setting of our numerical experiments, we prove that the running time of OAVI is $O(m\text{poly}(n))$ instead of $O(m^3\text{poly}(n))$, as previously shown in [9].

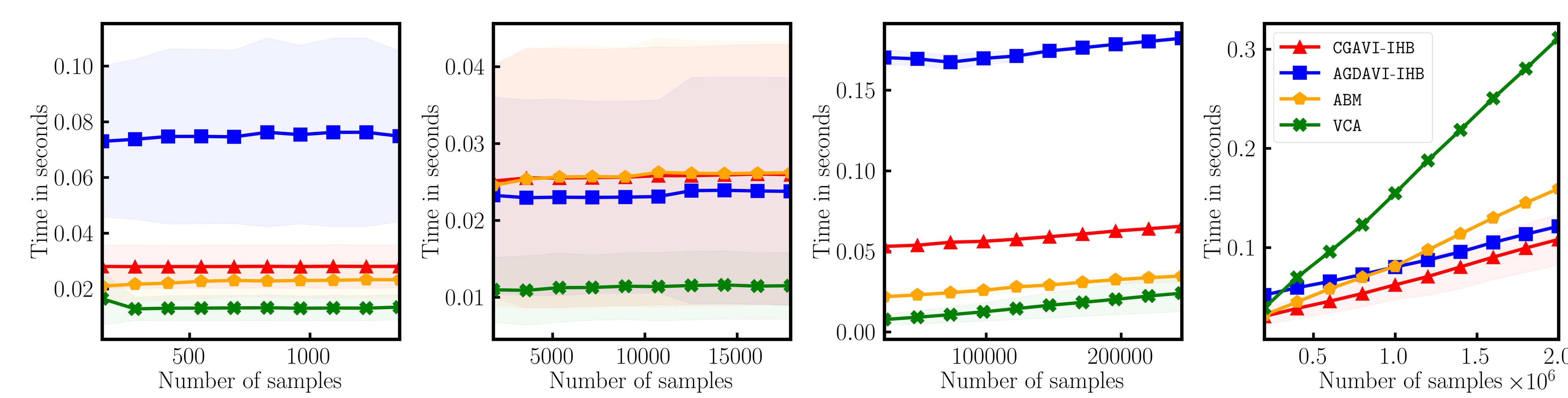
Experiment: inverse Hessian boosting (IHB) and weak inverse Hessian boosting (WIHB)



(a) Bank. (b) Htru. (c) Skin. (d) Synthetic.

Figure 1: Training time comparisons with fixed $\psi = 0.005$, averaged over ten random runs with shaded standard deviations. Inverse Hessian boosting (IHB) exploits inverse Hessian information to reduce the time required in OAVI to solve the convex subproblems. Weak inverse Hessian boosting (WIHB) is a sparsity-preserving variant of IHB. CGAVI-IHB is faster than BPCGAVI-WIHB, which is faster than BPCGAVI.

Experiment: comparison to the approximate Buchber-Möller algorithm (ABM) and vanishing component analysis (VCA)



(a) Bank. (b) Htru. (c) Skin. (d) Synthetic.

Figure 2: Training time comparisons, averaged over ten random runs with shaded standard deviations. For small data sets bank, htru, and skin, ABM and VCA are faster than OAVI, but for synthetic, the training times of ABM and VCA scale worse than OAVI's.

References

[1] David Cox, John Little, and Donal O'Shea. *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*. Springer Science & Business Media, 2013.

[2] Daniel Heldt, Martin Kreuzer, Sebastian Pokutta, and Hennie Poulisse. Approximate computation of zero-dimensional polynomial ideals. *Journal of Symbolic Computation*, 44(11):1566–1591, 2009.

[3] Simon Lacoste-Julien and Martin Jaggi. On the global linear convergence of Frank-Wolfe optimization variants. In *Proceedings of Advances in Neural Information Processing Systems*, pages 496–504, 2015.

[4] Jan Limbeck. Computation of approximate border bases and applications. 2013.

[5] Roi Livni, David Lehavi, Sagi Schein, Hila Nachliely, Shai Shalev-Shwartz, and Amir Globerson. Vanishing component analysis. In *Proceedings of the International Conference on Machine Learning*, pages 597–605. PMLR, 2013.

[6] H Michael Möller and Bruno Buchberger. The construction of multivariate polynomials with preassigned zeros. In *European Computer Algebra Conference*, pages 24–31. Springer, 1982.

[7] Kazuma K Tsuji, Ken'ichiro Tanaka, and Sebastian Pokutta. Pairwise conditional gradients without swap steps and sparser kernel herding. In *Proceedings of the International Conference on Machine Learning*, pages 21864–21883. PMLR, 2022.

[8] Elias Wirth, Hiroshi Kera, and Sebastian Pokutta. Approximate vanishing ideal computations at scale. In *Proceedings of the International Conference on Learning Representations*, 2023.

[9] Elias Wirth and Sebastian Pokutta. Conditional gradients for the approximate vanishing ideal. *arXiv preprint arXiv:2202.03349*, 2022.