Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 11, n. 1, 2025.

Trabalho apresentado no XLIII CNMAC, Centro de Convenções do Armação Resort - Porto de Galinhas - PE, 2024

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Machine learning topology of Calabi-Yau links

Tomás S. R. Silva[1], Henrique N. Sá Earp[2]
IMECC/UNICAMP, Campinas, SP

Abstract. Calabi-Yau links arise as special sphere fibrations over Calabi-Yau manifolds. In the 7-dimensional case, the links exhibit Sasakian and G_2 structures. In this summary, previous work is revisited, where machine learning and data science techniques are used to study topological quantities related to the Sasakian and G_2 geometries of contact Calabi-Yau 7-manifolds. Particularly, properties of the respective Gröbner bases are well-learnt, and machine learning of those further induced novel conjectures to be raised.

Keywords. G_2-manifolds, contact Calabi-Yau manifolds, machine learning, Hodge numbers, Crowley-Nördstrom invariant.

1 Introduction

Calabi-Yau manifolds have been a growing investigation topic lying in the interplay between geometry and theoretical physics since its introduction as a candidate for superstring compactification in [11]. Once the topology of Calabi-Yau compactification determines important properties of the effective field theory, the search for suitable manifolds that describe the observable universe has been an active research interest. However, the landscape of possible Calabi-Yau manifolds is enormous, expected to be at the order of 10^{10000} [5]. Thus, statistical and numerical methods provide means for feasibly extracting relevant data, of which machine learning techniques are becoming increasingly prominent. Notably, machine learning methods have been employed to predict Hodge numbers [15, 18], approximate Ricci-flat Calabi-Yau metrics [7, 14, 21], forecast line bundle cohomologies [20], generate new Calabi-Yau manifolds [8], and uncover volume bounds on Sasaki-Einstein manifolds [23].

As important as Calabi-Yau compactification is to string theory, 7-manifolds of holonomy G_2 are crucial to M-theory compactification [1, 2], which is manifestly 11-dimensional and requires compactification with a 7-dimensional manifold. This has motivated the investigation into physical applications of exceptional geometries with G_2 holonomy and the more general G_2-structure. In [3] machine learning techniques are used to investigate the topology of underlying G_2 and Sasakian structures of Calabi-Yau links arising from Calabi-Yau 3-folds defined as hypersurfaces in $\mathbb{P}^4(\mathbf{w})$ spaces.

In this summary, discussions and results obtained by the previous author's work [3] will be reviewed. In §2, we briefly discuss Calabi-Yau links and their topological invariants. In §3, machine learning techniques are used to investigate some of those topological invariants, including the Sasakian Hodge numbers and the Crowley-Nördstrom invariant. In §4, insights obtained via machine learning and data science are used to raise novel conjectures concerning Calabi-Yau links.

[1]tomas@ime.unicamp.br
[2]henrique.saearp@ime.unicamp.br

2 Calabi-Yau links

The structure group reductions on an odd-dimensional contact metric manifold (K^{2n+1}, η, ξ, g), where $\eta \in \Omega^1(K)$ denotes the contact form and $\xi \in \mathscr{X}(K)$ its (unit) dual Reeb field such that $\eta(\xi) = 1$; may be seen as even-dimensional structures transverse concerning a S^1-action along the fibres of a submersion $S^1 \to K \to V$. In particular, Sasakian geometry may be seen as transverse Kähler geometry, corresponding to the reduction of the transverse holonomy group to $\mathrm{U}(n)$. These are equipped with a transverse complex structure $J \in \mathrm{End}(TK)$ such that $J^2 = -\mathrm{I}_{TK} + \eta \otimes \xi$, yielding a decomposition of forms by basic bi-degree, and a transverse symplectic form $\omega = d\eta \in \Omega^{1,1}(K)$, all satisfying suitable compatibility conditions (see e.g. [25, §2] or [9, 10]). Furthermore, Sasakian manifolds with special transverse holonomy $\mathrm{SU}(n)$ are studied in [16, § 6.2.1].

Definition 2.1. *A Sasakian manifold $(K^{2n+1}, \eta, \xi, J, \Upsilon)$ is said to be a* ***contact Calabi–Yau manifold*** *(cCY) if Υ is a nowhere-vanishing transverse form of horizontal type $(n, 0)$, such that*

$$\Upsilon \wedge \bar{\Upsilon} = (-1)^{\frac{n(n+2)}{2}} \omega^n \quad \text{and} \quad d\Upsilon = 0, \quad \text{with} \quad \omega = d\eta.$$

Consider now a polynomial $f \in \mathbb{C}[z_1, \ldots, z_{n+2}]$, for $n \geq 2$. It is said to be **weighted homogeneous** of degree d with weight vector $\mathbf{w} = (w_1, \ldots, w_{n+2}) \in \mathbb{Z}^{n+2}_{>0}$, if it is homogeneous of order d with respect to the $\mathbb{C}^\times(\mathbf{w})$-action on $\mathbb{C}^{n+2}$, $(t, z) \mapsto t \cdot z = (t^{w_1} z_1, \ldots, t^{w_{n+2}} z_{n+2})$. Such an f defines an affine variety $V_f = (f) = \{z \in \mathbb{C}^{n+2} \mid f(z) = 0\}$, which in general admits the origin as a singular point.

Assuming that the origin is an isolated singularity, Milnor [24] showed that the intersection of V_f with a small hypersphere centred at the origin S^{2n+3}_ε is a compact smooth $(2n+1)$-manifold $K_f = V_f \cap S^{2n+3}_\varepsilon$, the so-called weighted link of the singularity. A weighted link K_f of degree d and weight w is a **Calabi-Yau link** if the following condition holds, which precisely guarantees the existence of a cCY structure on K_f:

$$d = \sum_{i=1}^{n+2} w_i, \tag{1}$$

Sasakian Hodge numbers of a CY link. The $\mathbb{C}^\times(\mathbf{w})$-action on $\mathbb{C}^{n+2}$ induces a contact-metric S^1-action on K_f. It admits finitely many distinct isotropy subgroups, contained in some finite subgroup $\Gamma \subset S^1$, so that K_f admits a double fibration over a projective n-orbifold $V \subset \mathbb{P}^{n+1}(\mathbf{w})$, $\pi : K_f \longrightarrow K_f/\Gamma \longrightarrow K_f/S^1 = (V_f \setminus \{0\})/\mathbb{C}^\times := V_f^*$. The following key theorem allows us to compute certain mixed Hodge numbers $h^{p,q}(K_f)$ from the dimensions of the primitive cohomology groups $H^n_0(V_f^*)$, for $p + q = n$, which in turn can be obtained from the Milnor algebra.

Theorem 2.1 ([19, Theorem 1.2], [26, 27])**.** *Let f be a $\mathbf{w}$-homogeneous polynomial on $\mathbb{C}^n$ of degree d. Given $p+q = n$, let $\ell = (p+1)d - \sum_i w_i$, and denote by $(\mathbb{M}_f)_\ell$ the linear subspace of the Milnor algebra consisting of degree ℓ elements.*

$$h^{p,q}(K_f) = \dim_{\mathbb{C}}(\mathbb{M}_f)_\ell.$$

When Equation (1) is satisfied, i.e. K_f is a Calabi-Yau link, the condition reduces to $\ell = pd$.

The Crowley-Nordström invariant on cCY 7-manifolds. For an arbitrary closed 7-manifold with G_2-structure (Y^7, φ), Crowley and Nordström have defined a $\mathbb{Z}/48\mathbb{Z}$-valued homotopy invariant $\nu(\varphi)$, which is a combination of topological data from a compact coboundary 8-manifold with $\mathrm{Spin}(7)$-structure (W^8, Ψ) extending (Y^7, φ), in the sense that $Y = \partial W$ and $\Psi \mid_Y = \varphi$, where χ the real Euler characteristic and σ is the signature:

$$\nu(\varphi) := \chi(W) - 3\sigma(W) \bmod 48, \tag{2}$$

G_2-structure on cCY 7-manifolds. Finally, specialising to real dimension 7 ($n = 3$), a contact Calabi–Yau structure naturally induces a coclosed G_2-structure (with symmetric torsion):

Proposition 2.1 ([16, Corollary 6.8])**.** *Every cCY manifold $(K^7, \eta, \xi, J, \Upsilon)$ is an S^1-bundle $\pi : K \to V$ over a Calabi–Yau 3-orbifold (V, ω, Υ), with connection 1-form η and curvature*

$$d\eta = \omega, \tag{3}$$

and it carries a cocalibrated G_2-structure

$$\varphi := \eta \wedge \omega + \mathrm{Re}\,\Upsilon, \tag{4}$$

with torsion $d\varphi = \omega \wedge \omega$ and Hodge dual 4-form

$$\psi = *\varphi = \frac{1}{2}\omega \wedge \omega + \eta \wedge \mathrm{Im}\,\Upsilon. \tag{5}$$

3 Topological invariants & machine learning

In [3], an exhaustive database of Calabi-Yau links from Calabi-Yau 3-folds defined as hypersurfaces in $\mathbb{P}^4(\mathbf{w})$ spaces was constructed[3]. The Calabi-Yau 3-folds arising in the link construction are hypersurfaces in complex weighted projective space $\mathbb{P}^4(\mathbf{w})$. Such spaces are compact Fano manifolds (with positive curvature), constructed through identification of $\mathbb{C}^5$ with a weight vector of 5 entries, i.e., $(z_0, \ldots, z_4) \sim (\lambda^{w_0} z_0, \ldots, \lambda^{w_4} z_4) \forall \lambda \in \mathbb{C}^*$. It was shown in [12], that the list of weight vector combinations leads to unique weighted projective spaces whose anticanonical divisors are compact and Ricci-flat is finite, with $N = 7555$ possible cases.

Therefore, for each of the 7555 possible links in the generated database, the respective Gröbner basis was computed, and invariants as Sasakian (i.e. transverse Kähler) Hodge numbers $\{h^{3,0}, h^{2,1}\}$ and Crowley-Nördstrom invariant ν were obtained via algorithmic implementation of Theorem 2.1 and Eq.(2). For all the links in the database, it was observed that $h^{3,0} = 1$, interestingly matching all the constituent Calabi-Yau manifolds used in their construction. The distributions of $h^{2,1}$'s and ν's can be visualised in Figures 1 and 2, respectively. Figure 3 shows the Calabi-Yau complex threefold $h^{2,1}$ values versus the Sasakian transverse $h^{2,1}_S$.

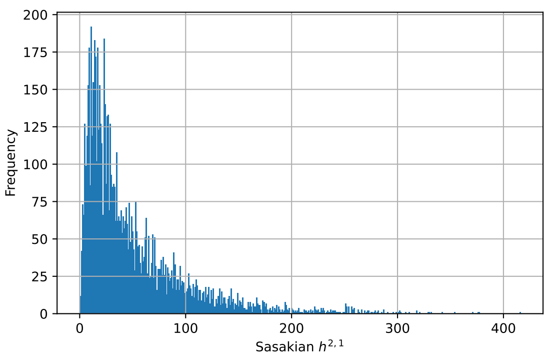

Figure 1: Distribution of $h^{2,1}_S$ [3].

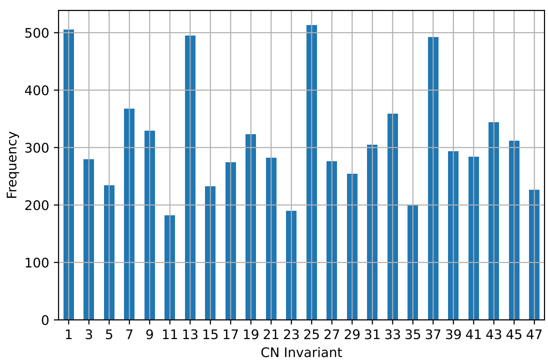

Figure 2: Distribution of ν [3].

[3] Data and code available at: https://github.com/TomasSilva/MLcCY7.git

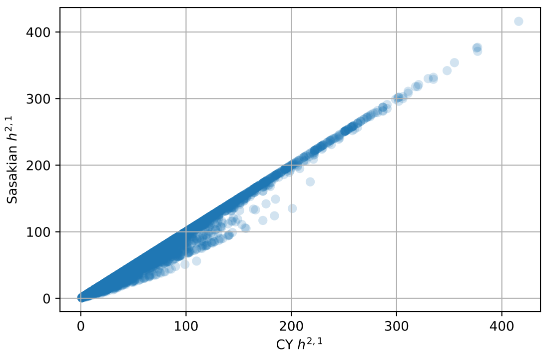

Figure 3: CY complex threefold $h^{2,1}$ values against the Sasakian transverse $h^{2,1}$ [3].

Once built the database of Calabi-Yau links, and motivated by previous works that successfully utilised supervised machine learning to predict Calabi-Yau Hodge numbers [15, 17, 18], similar techniques were employed in [3] to predict the computed link topological properties using only the weights defining the ambient spaces $\mathbb{P}^4(\mathbf{w})$. However, where in the Calabi-Yau case there is a known explicit formula mapping the weights to the respective Calabi-Yau Hodge numbers, there is no known direct relation bypassing the information of both the Calabi-Yau polynomial choice and the particularly expensive computation of the associated Gröbner basis used to compute the Sasakian Hodge numbers. Hence, machine learning prediction of Gröbner basis related properties can not only speedup expensive computations, but also give meaningful insights about those.

A field of machine learning is supervised learning, which exploits fitting techniques trained on pairs of input and output data. Given a dataset, divide it into training and testing subsets. Firstly, during a training stage, an optimiser method adjusts some architecture's parameters across batches of the training dataset to minimise some (loss) function of the true output and predicted output for each batch input. After rounds of the training process on the training dataset, the final trained architecture is then evaluated on its predictions for the independent testing dataset, allowing an evaluation of performance. The processes of training and testing may be repeated on multiple partitions of the original dataset on independent but identically designed architectures to provide means of averaging and calculating the confidence of the learning metrics (see e.g. [22]). The metrics utilised to evaluate the supervised learning process in [3] were the Mean Absolute Error (MSE) and the Coefficient of Determination (R^2), defined below:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad \in \quad (-\infty, 1]. \tag{7}$$

The MAE metric (6) computes the mean deviation of the predicted values from the true values. In turn, the R^2 metric (7) is the proportion of the variance in the dependent variable that is predictable from the independent variable(s), implying that a score close to 1 means the regression model is a good fit, whereas a score close to 0 means the model is a poor fit.

In [3], a supervised neural network (NN) (see e.g. [6]) and a symbolic regressor (SR) [13] were utilised to predict some of the CY link topological invariants discussed in §2 using each weight vector $\mathbf{w} = (w_i)_{i=0,\ldots,4}$ alone. Whilst the prediction of the Crowley-Nördstrom invariant didn't perform well, both methods perform surprisingly well predicting the Sasakian $h^{2,1}$ (see Table 1).

Method	R^2	**MAE**
NN	≈ 0.969	≈ 5.53
SR	≈ 0.99	≈ 2.6

Table 1: Supervised learning measures for $h_S^{2,1}$ prediction [3].

These results are equivalently strong and exemplify the efficacy of ML methods in predicting more subtle topological parameters. Figures 4 (resp. 5) illustrates the performance of NN (resp. SR) predictions against the true values of $h_S^{2,1}$ for each 7555 database link. Moreover, while the NN behaves as a black box oracle that predicts the $h_S^{2,1}$ for a giver weight vector (w_i), the SR produces interpretable relations between inputs and outputs, as in Eq.(8).

$$h^{2,1}_{\texttt{PySR}}(w_0,\ldots,w_4) = \frac{14.91 w_1 \left(w_0 w_4 + w_3 \left(w_0 + w_3\right)\right) + 10.02 w_2 w_3 \left(w_0 + w_4 + 0.77\right)}{w_0 w_1 w_2 w_3}. \tag{8}$$

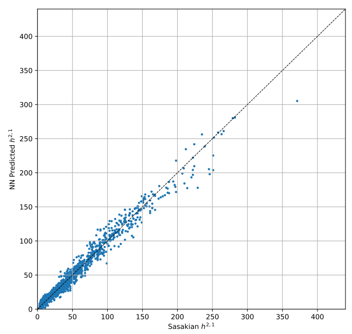

Figure 4: NN prediction vs. actual $h_S^{2,1}$ [3].

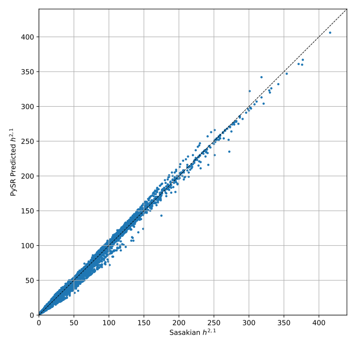

Figure 5: SR prediction vs. actual $h_S^{2,1}$ [3].

4 Novel conjectures

Computing the considered invariants for a variety of polynomials, with the correct singularity structure and defined from the same weight system, showed the same invariant values each time. In face of those results, the work summarised here [3] conjectured extension of the R-equivalence of [4], such that all polynomials with the same weight system would have isomorphic linear subspaces in their respective Milnor algebras.

Conjecture 4.1 ([3, Conjecture 7])**.** *Two weighted homogeneous polynomials f, g on $\mathbb{C}^n$ are said to be* ***weakly R-equivalent*** *if the respective ℓ-degree linear subspaces of their Milnor algebras are isomorphic, for each ℓ such that $p+q=n$, as in Theorem 2.1. Consider two weighted homogenous polynomials f, g on $\mathbb{C}^n$ of the same degree d; if their weight vectors $\mathbf{w}_f$ and $\mathbf{w}_g$ coincide (up to permutations), then f and g are weakly R-equivalent.*

Furthermore, one may also directly compare the values of the Sasakian Hodge number $h_S^{2,1}$ and the $h_{CY}^{2,1}$ of the Calabi-Yau 3-fold used in the construction, as illustrated in Figure 3. As shown in this figure, an unexpected upper bound is set by the CalabiYau 3-folds. This lead to another novel conjecture about the link construction method.

Conjecture 4.2 ([3, Conjecture 8]). *The Sasakian Hodge number $h_S^{2,1}$ for a Calabi-Yau link is bounded above by the Hodge number $h_{CY}^{2,1}$ of the Calabi-Yau 3-fold built from the same* **w**-*homogeneous polynomial:*

$$h_S^{2,1} \leq h_{CY}^{2,1} \ . \tag{9}$$

Future work is needed to prove these raised conjectures, expand the applications of machine learning to other Gröbner bases dependent properties, and use machine learning interpretable methods to help uncover the prospective formula for Calabi-Yau link Sasakian Hodge numbers from the weight information.

Acknowledgements

The authors would like to thank Daattavya Aggarwal, Yang-Hui He, Elli Heyes, and Edward Hirst for collaborating on the work summarized here. HSE is supported by São Paulo Research Foundation (FAPESP) grants 2018/21391-1, 2020/09838-0 and 2021/ 04065-6, the Brazilian National Council for Scientific and Technological Development (CNPq) grant 307217/2017-5, and UK Royal Society grant NMG\R1\191068. TSRS is supported by the São Paulo Research Foundation (FAPESP) grant 2022/09891-4.

References

[1] B. Acharya and S. Gukov. "M theory and singularities of exceptional holonomy manifolds". In: **Physics Reports** 392.3 (2004), pp. 121–189.

[2] B. Acharya and E. Witten. "Chiral fermions from manifolds of G_2 holonomy". In: (2001). eprint: `arXiv:hep-th/0109152v1`.

[3] Daattavya Aggarwal, Yang-Hui He, Elli Heyes, Edward Hirst, Henrique N. Sá Earp, and Tomás S.R. Silva. "Machine learning Sasakian and G2 topology on contact Calabi-Yau 7-manifolds". In: **Physics Letters B** 850 (2024), p. 138517. ISSN: 0370-2693. DOI: `10.1016/j.physletb.2024.138517`. URL: `https://www.sciencedirect.com/science/article/pii/S0370269324000753`.

[4] Imran Ahmed. "Weighted homogeneous polynomials with isomorphic Milnor algebras". In: **Journal of Prime Research in Mathematics** (2012).

[5] Ross Altman, Jonathan Carifio, James Halverson, and Brent D. Nelson. "Estimating Calabi-Yau hypersurface and triangulation counts with equation learners". In: **Journal of High Energy Physics** 2019.3 (Mar. 2019). ISSN: 1029-8479. DOI: `10.1007/jhep03(2019)186`. URL: `http://dx.doi.org/10.1007/JHEP03(2019)186`.

[6] James A Anderson. **An introduction to neural networks**. MIT press, 1995.

[7] Anthony Ashmore, Yang-Hui He, and Burt A. Ovrut. "Machine Learning Calab-Yau Metrics". In: **Fortschritte der Physik** 68.9 (Aug. 2020), p. 2000068. DOI: `10.1002/prop.202000068`.

[8] Per Berglund, Yang-Hui He, Elli Heyes, Edward Hirst, Vishnu Jejjala, and Andre Lukas. "New Calabi-Yau Manifolds from Genetic Algorithms". In: (June 2023). arXiv: `2306.06159 [hep-th]`.

[9] C. P. Boyer and K. Galicki. **Sasakian geometry**. Oxford Mathematical Monographs. Oxford University Press, Oxford, 2008, pp. xii+613. ISBN: 978-0-19-856495-9.

[10] O. Calvo-Andrade, L. Rodríguez, and H. N. Sá Earp. "Gauge theory and G_2-geometry on Calabi-Yau links". In: **Rev. Mat. Iberoam.** 36.6 (2020), pp. 1753–1778. eprint: `1606.09271[math.DG]`. URL: `https://www.ems-ph.org/journals/show_pdf.php?issn=0213-2230&vol=36&iss=6&rank=6`.

[11] P. Candelas, Gary T. Horowitz, Andrew Strominger, and Edward Witten. “Vacuum configurations for superstrings”. In: **Nuclear Physics B** 258 (C Jan. 1985), pp. 46–74. ISSN: 0550-3213. DOI: 10.1016/0550-3213(85)90602-9.

[12] P. Candelas, M. Lynker, and R. Schimmrigk. “Calabi-Yau manifolds in weighted P4”. In: **Nuclear Physics B** 341.2 (1990), pp. 383–402. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(90)90185-G. URL: https://www.sciencedirect.com/science/article/pii/055032139090185G.

[13] Miles Cranmer. **Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl**. 2023. arXiv: 2305.01582 [astro-ph.IM].

[14] Michael R. Douglas, Subramanian Lakshminarasimhan, and Yidi Qi. “Numerical Calabi-Yau metrics from holomorphic networks”. In: (Dec. 2020). arXiv: 2012.04797 [hep-th].

[15] Harold Erbin and Riccardo Finotello. “Machine learning for complete intersection Calabi-Yau manifolds: A methodological study”. In: **Physical Review D** 103.12 (June 2021). DOI: 10.1103/physrevd.103.126014.

[16] G. Habib and L. Vezzoni. “Some remarks on Calabi-Yau and hyper-Kähler foliations”. In: **Differential Geom. Appl.** 41 (2015), pp. 12–32. ISSN: 0926-2245. DOI: 10.1016/j.difgeo.2015.03.006.

[17] Yang-Hui He. “Deep-Learning the Landscape”. In: (June 2017). arXiv: 1706.02714 [hep-th].

[18] Yang-Hui He and Andre Lukas. “Machine Learning Calabi-Yau Four-folds”. In: **Phys. Lett. B** 815 (2021), p. 136139. DOI: 10.1016/j.physletb.2021.136139. arXiv: 2009.02544 [hep-th].

[19] Mitsuhiro Itoh. “SASAKIAN MANIFOLDS, HODGE DECOMPOSITION AND MILNOR ALGEBRAS”. In: **Kyushu J. Math** 58 (2004), pp. 121–140. URL: https://www.jstage.jst.go.jp/article/kyushujm/58/1/58_1_121/_article/-char/ja/.

[20] Daniel Klaewer and Lorenz Schlechter. “Machine learning line bundle cohomologies of hypersurfaces in toric varieties”. In: **Physics Letters B** 789 (Feb. 2019), pp. 438–443. DOI: 10.1016/j.physletb.2019.01.002.

[21] Magdalena Larfors, Andre Lukas, Fabian Ruehle, and Robin Schneider. “Learning Size and Shape of Calabi-Yau Spaces”. In: (Nov. 2021). arXiv: 2111.01436 [hep-th].

[22] Christopher M. **Pattern Recognition and Machine Learning**. en. 1st ed. Information Science and Statistics. New York, NY: Springer, Aug. 2006.

[23] Maksymilian Manko. “An Upper Bound on the Critical Volume in a Class of Toric Sasaki-Einstein Manifolds”. In: (Sept. 2022). arXiv: 2209.14029 [hep-th].

[24] John Milnor. **Singular points of complex hypersurfaces**. 1969.

[25] Luis E Portilla and Henrique N Sá Earp. “Instantons on Sasakian 7-manifolds”. In: **The Quarterly Journal of Mathematics** (2023). ISSN: 0033-5606. DOI: 10.1093/qmath/haad011. eprint: https://academic.oup.com/qjmath/advance-article-pdf/doi/10.1093/qmath/haad011/49696708/haad011.pdf.

[26] J. H. M. Steenbrink. **Mixed Hodge structures associated with isolated singularities**. 1983. DOI: 10.1090/PSPUM/040.2/713277.

[27] Joseph Steenbrink. “Intersection form for quasi-homogeneous singularities”. In: **Compositio Math.** 34.2 (1977), pp. 211–223. ISSN: 0010-437X.

