Accelerating inverse design using algorithmically obtained DFT gradients

Hochaktuell, Unkonventionalität mehr hervorheben!

1 Summary

1 page

The need for new materials is a common theme to major societal challenges of the 21st century, such as drug design, communication, energy production and storage, green chemistry. Proposed solutions are often fundamentally limited by the physical and chemical properties of available materials. Moreover as technologies become more specialised, the property requirements become more complex, too. Consider for example the case of transparent conductors in solar cells or touch-screen displays, which require an ideal material to satisfy the competing demands of high optical transparency and high electric conductivity [1]. The traditional direct design approach is based on incremental modification of known materials, followed by computational or experimental studies to check the properties of resulting candidates. Being highly dependent on human expertise this is not only hard to automatise, but also extremely error prone for the counterintuitive property combinations needed in such specific application regimes. The converse idea — called inverse design — is to perform a systematic search through a given design space in order to find those compounds, which match a set of desired properties [2]. In recent years this approach has become remarkably successful [3–8] leading for example to the discovery of novel semiconductors, electrocatalysts, materials for hydrogen storage or for Li-ion batteries. Despite these successes the available design spaces have not yet been exhaustively searched. One plain reason is their size. The number of possible metal alloys, for example, is just gigantic, since it scales exponentially with respect to the 90 metallic elements in the periodic table [6]. More efficient inverse design approaches, which are able to traverse larger design spaces, are thus desperately needed to accelerate the discovery of novel materials within the vast parts of chemical space we have missed so far.

On top of this as experimental techniques keep improving novel regimes for materials design are unlocked. A recent example is deep elastic strain engineering (ESE), rooted in the realisation of ultralarge tensile strains in diamond nanoneedles [9, 10]. This has lead to unexpected paths for reversible metallisation [11] or superconductivity [12] in diamond. Considering past works to alter the physical or chemical properties based on (way smaller) strains [9, 10, 13–18] and the billions of dollars of product value these works have created for silicon-based semiconductors [17], it is fair to say this has opened an exciting historic window of opportunity for stain-based inverse design. From the theoretical perspective initial works exploring this have just appeared, e.g. with the motivation to optimise diamond bandstructures for microelectronics application [10, 11, 19–21], but many exciting discoveries remain to be uncovered.

Modern inverse design approaches employ a variety of techniques including systematic screening, metaheuristics, evolutionary algorithms, statistical surrogates or machine-learned generative models to drive the design space search. At their heart these methods require first-principle property simulations, which can serve as training data or they are directly a part of the *objective function*, which is to be globally optimised over the design space [4, 5, 7]. The cost of these property simulations, typically employing density-functional theory, is the main bottleneck in inverse design. In the context of Bayesian optimisation — an established global optimisation strategy — recent mathematical work suggests that the objective function needs to be computed far fewer times if one also makes use of *objective function gradients* [22–27].

Importing this exciting development into materials design would require the computation of property gradients with respect to the design space parameters. For example in the strain-based inverse design of band structures we would need the gradient of the band energies with respect to the material strain. Since traditional DFT codes do not allow the computation of such gradients, this aspect is so far completely unexplored. However, recent advances with respect to algorithmic differentiation and in particular the appearance for algorithmically

differentiable codes for density-functional theory have now made such explorations feasible. citations

The main aim of this work is exactly to attempt the gradient-based design space exploration on the example of the strain-based optimisation of band structures. For this work we will build upon the algorithmically differentiable DFTK code for plane-wave density functional theory as well as Bayesian optimisation methods.

2 Project plan

2.1 State of the art

2.1.1 The optimisation problem underlying inverse design

The goal of inverse design is to find materials from a given design space, which best meet a desired objective. For example considering the strain-based engineering of the diamond band structure, our design parameter λ is the strain realised in the diamond crystal and the objective O is some way of measuring how the band gap or band curvatures deviates from our desired behaviour. Beyond the implied strain λ, the system is free to relax e.g. its atomic positions. These additional degrees of freedom we collect in R. For each instance of the parameters λ and R we can employ a first-principles method to compute the total energy $E(\lambda, R)$ as well as all other quantities required to obtain the objective $O(\lambda, R)$. Since our interest is to discover materials, which can be synthesised and stably employed in practice, we cannot just minimise $O(\lambda, R)$ over the design space. Additionally need to ensure that we only consider minimal-energy structures. This leads to the *bilevel optimisation problem*

$$\begin{aligned} &\min_{\lambda,R} O(\lambda, R) \\ \text{subject to} \quad &\min_{R} E(\lambda, R), \end{aligned} \tag{1}$$

where the upper-level (UL) optimisation problem of inverse design is constrained by a lower-level (LL) energy minimisation problem. A common technique to deal with (1) is *single-level reduction* [28, 29]: The stationarity condition of the LL problem, i.e. achieving zero force,

$$0 = F(\lambda, R) = \left(\frac{\partial E}{\partial R}\right)(\lambda, R) \tag{2}$$

defines an implicit function $R_*(\lambda)$ around a local minimum of E wrt. R. Assuming the lower-level problem to be convex $R_*(\lambda)$ is valid on the whole domain of λ and we can insert it into the UL objective

$$\widetilde{O}(\lambda) = O(\lambda, R_*(\lambda)). \tag{3}$$

We are overall faced with the standard minimisation problem $\min_\lambda \widetilde{O}(\lambda)$. For solving such inverse materials design problems a great deal of methods have been proposed in the past [4, 5, 7], e.g. systematic screening, evolutionary algorithms, surrogatisation techniques or generative models. Most (if not all) assume LL convexity and combine a local structure relaxation with a global search over the design space. However, energy minimisation problems are not convex in general. Given sufficient flexibility in the structural parameters R, the LL problem could thus converge to a metastable structure, which in turn could misguide the design space search. Within the context of strain-based inverse design this could, for example, occur when non-isotropic strains are applied to surfaces or layered materials, provoking a symmetry breaking. The issue of non-convex LL problems has recently received considerable attention in the bilevel optimisation community. A number of tailored algorithms have been proposed [30–34], which are promising to reliably expand the scope of inverse design to more involved settings.

2.1.2 Gradient-accelerated Bayesian optimisation techniques

A popular surrogatisation technique for global optimisation is Bayesian Optimisation (BO) [35, 36]. In these methods a small number of queries of the objective function $\widetilde{O}(\lambda)$ is employed to construct a Gaussian Process (GP) model. Based on the built-in uncertainty quantification capabilities of the GP, the *acquisition function* automatically chooses further points, where the objective function should be evaluated. A variety of acquisition functions can be found in the literature, which vary amongst others in their trade-off between exploration

(sample where the GP knows little) and exploitation (sample where the GP suspects the global minimiser). BO methods have previously been applied both in the context of bilevel problems [37–39] as well as materials design [40, 41]. With this work we want to explore in particular recent ideas, which suggests that BO techniques need far fewer queries from the objective if also objective function gradients are employed [22–27]. Assuming that these objective gradients can be computed with sufficiently little overhead over computing the objective value alone, this overall reduces the cost of the global optimisation.

To employ gradient-accelerated optimisation techniques we need the derivative of the objective function $\widetilde{O}$. For a wide range of inverse design problems this function is either differentiable or can be made differentiable by minor adjustments (e.g. appropriate smoothening). Under slight abuse of notation we obtain

$$\frac{d\widetilde{O}}{d\lambda} = \frac{dO}{d\lambda} = \frac{\partial O}{\partial \lambda} + \frac{\partial O}{\partial R}\frac{\partial R_*}{\partial \lambda}, \tag{4}$$

where we used the chain rule as well as the function R_* implicitly defined by the LL energy minimisation problem. By the implicit function theorem we can obtain the partial derivative of R_* as

$$\frac{\partial R_*}{\partial \lambda} = -\left(\frac{\partial F}{\partial R}\right)^{-1}\frac{\partial F}{\partial \lambda} = -\left(\frac{\partial^2 E}{\partial^2 R}\right)^{-1}\frac{\partial^2 E}{\partial R \partial \lambda}. \tag{5}$$

For our simulations in this work we will focus on density-functional theory (DFT), which is widely employed for inverse design due to its favourable balance between cost and accuracy. Computing $E(\lambda, R)$ and $O(\lambda, R)$ thus implies that we first solve the self-consistent field problem (SCF) and use the self-consistent density (and orbitals) to compute E or O. Similarly to compute the gradient (4) the DFT code needs to be able to deliver all partial derivatives of (4) and (5). In turn this depends on the self-consistent density, but also its response to the perturbations induced by changes in R or λ, which need to be computed via density-functional perturbation theory (DFPT).

Considering gradient-accelerated BO this implies two challenges. Firstly DFPT is associated with a non-negligible cost, which implies that the gradient (4) is not cheap either. However, this cost is to be compared with the cost of evaluating $\widetilde{O}$, which requires the solution of two nested iterative processes (SCF and structure relaxation). The hypothesis underlying this work is thus that the overhead of computing (4) will indeed be compensated by using far fewer queries in the gradient-accelerated BO procedure. Secondly, providing all required partial derivatives is a considerable implementation effort, in particular since the objective function $O(\lambda, R)$ may encompass multiple DFT quantities (depending on the target of the design-space search) and thus all their derivatives would be required. Additionally the respective derivatives with respect to λ or R may not be related to physical quantities, i.e. not necessarily of much use beyond inverse design. Overall a manual implementation is thus not an economical use of human time.

2.1.3 Algorithmic differentiation in materials modelling

An alternative to manually implementing the derivatives in (4) into a DFT code provides algorithmic differentiation (AD), also referred to as automatic differentiation or differentiable programming. The roots of AD date back to the early 1960s [42]. Its application to artificial intelligence and training of neural networks (a special case of numerical simulation software) has been investigated since about the same time, see [43]. By now AD is well established, see [44, 45] for a comprehensive coverage of AD and the extensive bibliography at `www.autodiff.org` for an overview of the technical challenges and successful applications. Fueled from the enormous success of neural-network approaches and the wide availability of AD-driven machine learning packages such as TensorFlow [46] or PyTorch [47] the application of AD has recently also gained considerable attraction in atomistic modelling [48, 49]. Promising flagship applications from the inverse design perspective

include (a) the generation of interatomic potentials from known folding patterns of proteins [50–52] or (b) the learning of DFT functionals from predictions at higher levels of theory or experimental data [53–56]. The latter work is closely linked to a number of differentiable DFT packages, which have recently appeared for Gaussian basis sets [56–58].

However, this project focuses on the plane-wave DFT setting, which is more conventional for the solid-state physics applications we target. In this regime AD techniques are far less developed due to methodological obstacles posed by the larger basis size. To the best of my knowledge only a single code with AD capabilities has been made available to date, namely the density-functional toolkit (DFTK) [59]. DFTK is a recent Julia-based [60] package, which in only 8000 lines of code already supports a good range of standard models (norm-conserving pseudopotentials, meta-GGA functionals) as well as calculations on systems of around 1000 electrons. On the AD side Julia brings an ecosystem of extremely powerful tools [61–63], which have in the past been applied to differentiate large software stacks [64] including GPU-enabled or MPI-parallelised code [65, 66]. So far DFTK only supports forward-mode AD [67], which is applicable to computing derivatives with respect to a small number of parameters — like the derivatives of $\widetilde{O}(\lambda)$ wrt. the 6 strain parameters we require. Under the hood the implemented expression of $\widetilde{O}$ as well as techniques such as the chain rule (e.g. equation (4)) or implicit differentiation (e.g. equation (5)) are utilised to build (and compile) an expression for the desired derivative. The latter is based on standard primitives such as DFPT problems, which DFTK solves to yield the derivative value. Of note the overall machinery is highly flexible wrt. the functional form of $\widetilde{O}$ — with AD automatically taking care of the bookkeeping which DFPT problems to invoke and how to combine their results. In its current stage the AD implementation in DFTK is functional, but so far works demonstrating its performance and suitability for applications are lacking. Since the computational bottlenecks (the DFPT problems) are clear this work strives to iron out the remaining practicalities (tolerances, initial guesses, ...) and to provide a first demonstration of AD-based gradients for plane-wave DFT in the context of inverse design.

2.2 Detailed description

2.2.1 Project goals

The high-level aim of this project is to investigate the perspectives for inverse design opened up by the recent integration of algorithmic differentiation (AD) into state-of-the-art DFT codes such as DFTK. We will guide and test our developments with reference to a highly relevant exemplary problem, namely the inverse design of band structures based on elastic strain engineering (ESE) — which we will refer to as our **focus problem**. This field has recently seen considerable experimental advances and the opportunities of strain-based property engineering have not yet been fully exploited. Two methodological goals will be targeted:

Efficient gradient-accelerated Bayesian optimisation (goal G1). Making use of the AD capabilities of DFTK a gradient-accelerated BO approach for inverse design will be developed, which allows a flexible selection of objective function and DFT functional. Based on the **focus problem** its computational cost will be benchmarked, in particular to understand if the additional gradient information is worth its computational cost.

Bilevel optimisation methods without lower-level (LL) convexity (goal G2). Numerous inverse design approaches rely on the LL structure relaxation problem to be convex, which in general is not the case. With respect to our **focus problem** we want to understand to what extend this can become an issue in practice. Moreover we will develop a general inverse design (applicable beyond the **focus problem**), which is suitable in the setting without LL convexity (LLC).

2.2.2 Work plan

To achieve our ambitious goals the project is laid out for 12 months and will proceed along the following three work packages (WP). Note that only personal costs are requested; computational time is covered by other funds.

WP1: Inverse design using quasi-Newton methods (4 months). The purpose of this WP is twofold. Firstly to develop an implementation for the computation of both $\widetilde{O}$ and its gradient wrt. λ (equations (4) and (5)) making use of DFTK and the AD packages of the Julia ecosystem. Secondly to employ this implementation to directly drive a quasi-Newton-based design space search towards a local minimum of $\min_\lambda \widetilde{O}(\lambda)$. The difficulty in this WP originates from the three layers of nested iterative solvers one needs to differentiate through (SCF, structure relaxation, objective minimisation). While AD methods promise to help with the overall bookkeeping in such situations, one is still faced with a number of practical complications to obtain an efficient algorithm. This includes aspects such as an appropriate adaptation of tolerances or the transfer of suitable initial guesses in the resulting nested linear problems, which are needed as part of the AD procedure (e.g. respective density-functional perturbation theory (DFPT) problems). To guide developments we will apply the quasi-Newton approach to our **focus problem**, e.g. to use it to find strains in diamond, which lead to a zero band gap, a direct band gap or which induce a phase transition to a material with Dirac cones (maybe leading to graphene). **Deliverables:** Efficient AD-based gradient computation of $\widetilde{O}$.

WP2: Gradient-accelerated BO-based inverse design (4 months). This WP focuses on employing the gradient information on $\widetilde{O}$ made available by **WP1** to integrate with recent work on gradient-driven BO methods [22–27]. The main challenge will be to balance the increased computational cost for obtaining the gradient information with a sufficient reduction on the number of required queries. Particularly promising in this directions are two approaches. Firstly, to use the gradient information not only for training the Gaussian Process (GP), but also within specific gradient-aware acquisition functions [25] to improve the quality of selected query points. Secondly, to exploit the noise-tolerance of GPs, more specifically that the GP surrogates underlying BO are able to support *different* noise levels in objective and gradient [24]. This enables us to reduce the number of iterations in the DFPT solvers for the gradient computation at the expense of having more noisy gradients. We will implement our ideas by interfacing DFTK with the existing BayesianOptimization [68] Julia package and rely on our **focus problem** for benchmarking our ideas. In particular we will consider the band structure design of silicon and diamond in order to compare our work to previous approaches based on neural networks [20]. **Deliverables: Goal G1**; publication evaluating and applying the method to the **focus problem**.

WP3: Investigating the case of non-convex LL problems (4 month). For general inverse design settings the convexity assumption in the LL structure relaxation is not justified. However, LL convexity may still be given if systems are small and highly symmetric — like the settings in which our **focus problem** has so far been investigated [10–12, 19–21]. Our goal in this WP is to push for more complex systems, in an attempt to explicitly find a counterexample, i.e. we will seek a material under strain λ, such that the structure relaxation $\min_R E(\lambda, R)$ has multiple local minimisers. We will use standard metaheuristics techniques (such as particle swarm methods [69]) to find the multiple minima of supercells or layers of main-group semiconductors, which are put under a large anisotropic strains to provoke symmetry breaking. If counterexamples cannot be easily found, this gives reassurance about assuming convexity in this setting. If they can be found, we have a good set of test cases for the second part of the WP, namely the development of inverse design methods without LL convexity. For doing so the recent primal-dual bilevel optimiser (PDBO) [34] method form the bilevel optimisation literature is promising, since it only requires minor modifications (in the form of additional constraints) to (1) to (3). The required gradients will be straightforward to implement building on top of the AD-based quasi-Newton method of **WP1**.

Deliverables: Goal G2; publication discussing LL convexity and applications of PDBO to inverse design.

2.2.3 Risk analysis and risk management

The goals aspired in this project are ambitious and aim to apply recent advances from algorithmic differentiation (AD) as well as gradient-accelerated Bayesian optimisation (BO) to inverse design problems for the first time. Both tools are furthermore hardly employed in materials modelling, such that the sketched research is associated with considerable risks. Below key risks are sketched along with opportunities in case of failure.

- **Gradient computation using AD is too inefficient for our needs:** The recent excitement in the Gaussian basis community around the use of AD for electronic structure methods and the flurry of respective AD-enabled software [56–58] illustrates the importance to make AD methods viable in the plane-wave context as well. Even if AD methods cannot be sufficiently improved to enable **WP2** and **WP3** I am convinced that all improvements of **WP1** will strengthen AD for plane-wave DFT and failure will naturally lead to relevant follow-up research to improve AD methods.
- **Gradient cost not offset by reducing number of queries in gradient-accelerated BO:** In recent years gradient-based methods have been proposed for numerous global optimisation approaches, not just BO [70–73]. In this sense the AD-based objective gradients of **WP1** add a completely new tool for inverse design, where **WP2** and **WP3** only provide two examples for its exploration. A failure in these two WPs will help to understand the limits of AD-based gradients and thus guide further exploration.
- **LL convexity is always given for our focus problem:** If finding a counterexample in **WP3** turns out to be prohibitively hard, a formal proof establishing this property (for the setting of **focus problem**) could be worth investigating. In this case the insight obtained in **WP3** will be value to formalising assumptions as well as the proof.

2.3 Impact

2.3. Beschreibung der potenziellen Wirkung («impact») des Forschungsprojekts;
* Inwiefern das Projekt das Potenzial hat, eine transformative Veränderung eines zentralen wis- senschaftlichen Themas herbeizuführen und/oder den Weg für ein neues Forschungsgebiet oder eine neue Methode / Technologie zu ebnen und/oder das Handeln und Denken in der Gesellschaft zu verändern.

A central element of all proposed work packages is to take recent ideas from mathematics and computer science (algorithmic differentiation (AD), Bayesian optimisation, bilevel optimisation) and apply them for the first time to inverse design problems. All development is directly guided by a highly impactful focus topic, the strain-based engineering of semiconductors, where experimental advances have recently opened novel opportunities for materials design. The project thus provides a bridge between numerous fundamental research topics and real-world applications, opening up further in-depth exploration in future endeavours.

Of these let me highlight the opportunities of algorithmic differentiation (AD) for plane-wave DFT. In this project AD is the central ingredient, which enables as the key novelty of this work to consider optimisation methods requiring objective gradients. Essentially the role of AD is to take care of setting up and solving the required DFPT problems and finally collecting the results to make up exactly the gradient corresponding to the user-defined objective function. Of note AD is in theory even more general than that and should allow the computation of *any* derivative of DFT output quantity (bands, density of states, density, ...) versus *any* DFT input quantity (atomic positions, lattice, DFT model...). If this this was fully available in a plane-wave DFT context this would allow any user (including novices to DFT development) to start computing their own DFT derivatives — even including derivatives no-one ever computed before them. For applications in inverse design of materials, but also inverse design of DFT models themselves [54] this brings a fundamentally new tool to the

table, which fits well into the theme of recently appearing gradient-driven global optimisation methods. Albeit the foundations for these methods are now available in a first plane-wave DFT code, considerable practical challenges remain. With this project we hope to contribute substantially to advancing such methods with one clear use case in mind. However, we are convinced our work can show their success in practice and thus not only provide fundamentally new methods for inverse materials design, but also a convincing first showcase for AD in plane-wave DFT.

Check and cleanup references

References

[1] L. Zhang, Y. Zhou, L. Guo, W. Zhao, A. Barnes, H.-T. Zhang, C. Eaton, Y. Zheng, M. Brahlek, H. F. Haneef, N. J. Podraza, M. H. W. Chan, V. Gopalan, K. M. Rabe and R. Engel-Herbert. Nature Materials, **15**, 204 (2016).

[2] A. Franceschetti and A. Zunger. Nature, **402**, 60 (1999).

[3] A. Jain, Y. Shin and K. A. Persson. Nat. Rev. Mater., **1**, 15004 (2016).

[4] B. Sanchez-Lengeling and A. Aspuru-Guzik. Science, **361**, 360 (2018).

[5] A. Zunger. Nature Reviews Chemistry, **2** (2018).

[6] K. Alberi, M. B. Nardelli, A. Zakutayev et al. J. Phys. D: Appl. Phys., **52**, 013001 (2019).

[7] J. Noh, G. H. Gu, S. Kim and Y. Jung. Chemical Science, **11**, 4871 (2020).

[8] S. Luo, T. Li, X. Wang et al. Wiley Interdiscip. Rev. Comput. Mol. Sci., **11**, e1489 (2021).

[9] A. Banerjee, D. Bernoulli, H. Zhang, M.-F. Yuen, J. Liu, J. Dong, F. Ding, J. Lu, M. Dao, W. Zhang, Y. Lu and S. Suresh. Science, **360**, 300 (2018).

[10] A. Nie, Y. Bu, P. Li, Y. Zhang, T. Jin, J. Liu, Z. Su, Y. Wang, J. He, Z. Liu, H. Wang, Y. Tian and W. Yang. Nature Communications, **10** (2019).

[11] Z. Shi, M. Dao, E. Tsymbalov, A. Shapeev, J. Li and S. Suresh. Proceedings of the National Academy of Sciences, **117**, 24634 (2020).

[12] C. Liu, X. Song, Q. Li, Y. Ma and C. Chen. Physical Review Letters, **124** (2020).

[13] H. M. Manasevit, I. S. Gergis and A. B. Jones. Applied Physics Letters, **41**, 464 (1982).

[14] Y. Sun, S. E. Thompson and T. Nishida. Journal of Applied Physics, **101**, 104503 (2007).

[15] L. Thulin and J. Guerra. Physical Review B, **77** (2008).

[16] S. A. Kukushkin and A. V. Osipov. Journal of Applied Physics, **113**, 024909 (2013).

[17] A. Lu, L. Yang, C. Dang, H. Wang, Y. Zhang, X. Li, H. Zhang and Y. Lu. Functional Diamond, **2**, 151 (2022).

[18] C. Dang, A. Lu, H. Wang, H. Zhang and Y. Lu. Journal of Semiconductors, **43**, 021801 (2022).

[19] Z. Shi, E. Tsymbalov, M. Dao, S. Suresh, A. Shapeev and J. Li. Proceedings of the National Academy of Sciences, **116**, 4117 (2019).

[20] E. Tsymbalov, Z. Shi, M. Dao, S. Suresh, J. Li and A. Shapeev. npj Computational Materials, **7** (2021).

[21] C. C. Price, A. Singh, N. C. Frey and V. B. Shenoy. Science Advances, **8** (2022).

[22] D. J. Lizotte. phdthesis, University of Alberta (2008).

[23] O.-P. Koistinen, E. Maras, A. Vehtari and H. Jónsson. Nanosystems: Physics, Chemistry, Mathematics, 925–935 (2016).

[24] D. Eriksson, K. Dong, E. Lee, D. Bindel and A. G. Wilson. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, vol. 31. Curran Associates, Inc. (2018).

[25] J. Wu, M. Poloczek, A. G. Wilson and P. Frazier. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc. (2017).

[26] S. Shekhar and T. Javidi. In A. Banerjee and K. Fukumizu (Eds.), *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, vol. 130 of *Proceedings of Machine Learning Research*, 2836–2844. PMLR (2021).

[27] S. E. Ament and C. P. Gomes. In K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu and S. Sabato (Eds.), *Proceedings of the 39th International Conference on Machine Learning*, vol. 162 of *Proceedings of Machine Learning Research*, 500–516. PMLR (2022).

[28] B. Colson, P. Marcotte and G. Savard. Annals of Operations Research, **153**, 235 (2007).

[29] A. Sinha, P. Malo and K. Deb. IEEE Transactions on Evolutionary Computation, **22**, 276 (2018).

[30] R. Liu, P. Mu, X. Yuan, S. Zeng and J. Zhang. In H. D. III and A. Singh (Eds.), *Proceedings of the 37th International Conference on Machine Learning*, vol. 119 of *Proceedings of Machine Learning Research*, 6305–6315. PMLR (2020).

[31] R. Liu, Y. Liu, S. Zeng and J. Zhang. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang and J. W. Vaughan (Eds.), *Advances in Neural Information Processing Systems*, vol. 34, 8662–8675. Curran Associates, Inc. (2021).

[32] H. Shen, Q. Xiao and T. Chen (2023).

[33] L. Chen, J. Xu and J. Zhang (2023).

[34] D. Sow, K. Ji, Z. Guan and Y. Liang (2022).

[35] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams and N. de Freitas. Proceedings of the IEEE, **104**, 148 (2016).

[36] P. I. Frazier (2018).

[37] E. Kieffer, G. Danoy, P. Bouvry and A. Nagih. In *GECCO '17: Genetic and Evolutionary Computation Conference*. ACM (2017).

[38] E. Kieffer, G. Danoy, M. R. Brust, P. Bouvry and A. Nagih. IEEE Transactions on Evolutionary Computation, **24**, 44 (2020).

[39] V. Dogan and S. Prestwich. *Bayesian Optimization with Multi-objective Acquisition Function for Bilevel Problems*, 409–422. Springer Nature Switzerland (2023).

[40] P. M. Tagade, S. P. Adiga, S. Pandian, M. S. Park, K. S. Hariharan and S. M. Kolake. npj Computational Materials, **5** (2019).

[41] R. Pollice, G. dos Passos Gomes, M. Aldeghi, R. J. Hickman, M. Krenn, C. Lavigne, M. Lindner-D'Addario, A. Nigam, C. T. Ser, Z. Yao and A. Aspuru-Guzik. Accounts of Chemical Research, **54**, 849 (2021).

[42] R. Wengert. Commun. ACM, **7**, 463 (1964).

[43] P. Werbos. *The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting*. Wiley (1994).

[44] A. Griewank and A. Walther. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. No. 105 in Other Titles in Applied Mathematics. SIAM, 2nd edn. (2008).

[45] U. Naumann. *The Art of Differentiating Computer Programs. An Introduction to Algorithmic Differentiation.* No. SE24 in Software, Environments, and Tools. SIAM (2012).

[46] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, R. Jozefowicz, Y. Jia, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, M. Schuster, R. Monga, S. Moore, D. Murray, C. Olah, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng. *TensorFlow, Large-scale machine learning on heterogeneous systems* (2015).

[47] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai and S. Chintala. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox and R. Garnett (Eds.), *Advances in Neural Information Processing Systems 32*, 8024–8035. Curran Associates, Inc. (2019).

[48] B. Ramsundar, D. Krishnamurthy and V. Viswanathan (2021).

[49] R. Pederson, B. Kalita and K. Burke. Nature Reviews Physics, **4**, 357 (2022).

[50] W. Wang, S. Axelrod and R. Gómez-Bombarelli (2020).

[51] J. G. Greener and D. T. Jones. PLOS ONE, **16**, e0256990 (2021).

[52] D. Schwalbe-Koda, A. R. Tan and R. Gómez-Bombarelli. Nature Communications, **12** (2021).

[53] L. Li, S. Hoyer, R. Pederson et al. Phys. Rev. Lett., **126**, 036401 (2021).

[54] J. Kirkpatrick, B. McMorrow, D. H. P. Turban, A. L. Gaunt, J. S. Spencer, A. G. D. G. Matthews, A. Obika, L. Thiry, M. Fortunato, D. Pfau, L. R. Castellanos, S. Petersen, A. W. R. Nelson, P. Kohli, P. Mori-Sánchez, D. Hassabis and A. J. Cohen. Science, **374**, 1385 (2021).

[55] M. Kasim and S. Vinko. Physical Review Letters, **127** (2021).

[56] Y. Chen, L. Zhang, H. Wang and W. E. Journal of Chemical Theory and Computation, **17**, 170 (2021).

[57] M. F. Kasim, S. Lehtola and S. M. Vinko. The Journal of Chemical Physics, **156**, 084801 (2022).

[58] X. Zhang and G. K.-L. Chan. The Journal of Chemical Physics, **157**, 204801 (2022).

[59] M. F. Herbst, A. Levitt and E. Cancès. Proc. of the JuliaCon Conf., **3**, 69 (2021).

[60] J. Bezanson, A. Edelman, S. Karpinski et al. SIAM Rev., **59**, 65 (2017).

[61] W. Moses and V. Churavy. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan and H. Lin (Eds.), *Advances in Neural Information Processing Systems*, vol. 33, 12472–12485. Curran Associates, Inc. (2020).

[62] J. Revels, M. Lubin and T. Papamarkou. *Forward-Mode Automatic Differentiation in Julia* (2016). Arxiv:1607.07892.

[63] M. Innes. *Don't Unroll Adjoint: Differentiating SSA-Form Programs* (2018). Arxiv:1810.07951.

[64] C. Rackauckas and Q. Nie. J. Open Research Softw., **5** (2017).

[65] W. S. Moses, V. Churavy, L. Paehler, J. Hückelheim, S. H. K. Narayanan, M. Schanen and J. Doerfert. In *SC '21: The International Conference for High Performance Computing, Networking, Storage and Analysis*. ACM (2021).

[66] W. S. Moses, S. H. K. Narayanan, L. Paehler, V. Churavy, M. Schanen, J. Hückelheim, J. Doerfert and P. Hovland. In *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*, SC '22. IEEE Press (2022).

[67] *DFTK features documentation*. `https://docs.dftk.org/stable/features/` (). Accessed: 2023-04-23.

[68] *BayesianOptimization.jl: Bayesian optimization for Julia* . `https://github.com/jbrea/BayesianOptimization.jl` (). Accessed: 2023-04-23.

[69] M. R. Bonyadi and Z. Michalewicz. Evolutionary Computation, **25**, 1 (2017).

[70] N. Noman and H. Iba. IEEE Transactions on Evolutionary Computation, **12**, 107 (2008).

[71] V. Plevris and M. Papadrakakis. Computer-Aided Civil and Infrastructure Engineering (2010).

[72] M. Fischetti and M. Stringher. *Embedding Simulated Annealing within Stochastic Gradient Descent*, 3–13. Springer International Publishing (2021).

[73] M. EL-Alem, A. Aboutahoun and S. Mahdi. Soft Computing, **25**, 2325 (2021).