

RESEARCH ARTICLE OPEN ACCESS

Score Matching for Differential Abundance Testing of Compositional High-Throughput Sequencing Data

Johannes Ostner^{1,2}  | Hongzhe Li³ | Christian L. Müller^{1,2,4}

¹Computational Health Center, Helmholtz Munich, Neuherberg, Germany | ²Institut für Statistik, Ludwig-Maximilians-Universität München, Munich, Germany | ³Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA | ⁴Center for Computational Mathematics, Flatiron Institute, New York, New York, USA

Correspondence: Johannes Ostner (johannes.ostner@online.de)

Received: 1 August 2025 | **Revised:** 21 March 2026 | **Accepted:** 25 March 2026

Keywords: compositional data | differential abundance | generative model | microbiome | score matching | single-cell RNA sequencing

ABSTRACT

The class of α - β power interaction models, proposed by [1], provides a general framework for modeling sparse compositional data with pairwise feature interactions. This class includes many distributions as special cases and enables modeling of zero entries through power transformations, making it particularly suitable for modern high-throughput sequencing data with excess zeros, including single-cell RNA-Seq and microbial amplicon data. Here, we present an extension of this class of models that allows inclusion of covariate information, thus enabling accurate characterization of covariate dependencies in heterogeneous populations. Combining this model with a tailored differential abundance (DA) test leads to a novel DA testing scheme, `cosmoDA`, that can reduce the false positive detection rate caused by correlated features. `cosmoDA` uses penalized generalized score matching for parsimonious model fitting. We show on simulated benchmarks that `cosmoDA` can accurately estimate feature interactions in the presence of population heterogeneity and significantly reduces the false discovery rate when testing for differential abundance of correlated features. Using single-cell and amplicon data, we illustrate `cosmoDA`'s ability to estimate data-adaptive Box–Cox-type data transformations and assess the impact of zero replacement and power transformations on downstream differential abundance results. `cosmoDA` is available at <https://github.com/bio-datascience/cosmoDA>.

1 | Introduction

Abundance matrices, detailing the compositional makeup of cellular constituents in a sample, are an important data modality derived from modern high-throughput sequencing (HTS) experiments, including amplicon sequencing [1, 2] and single-cell RNA-Sequencing (scRNA-Seq) [3, 4]. These matrices commonly have the form $\tilde{\mathbf{X}} \in \mathbb{N}_0^{n \times p}$ and show the abundance of p features (cell types or microbial taxa) in n tissues [5], bacterial communities [6], or microbial habitats [7]. Because the sequencing capacity of HTS experiments is technically limited, each sample only represents a small part of a larger population, rendering the sum of counts in a row non-quantitative and making the data

compositional [8]. Dividing each sample by its total sum yields relative abundance data, which is proportionally equivalent to the original data and constrained to the $(p - 1)$ -dimensional probability simplex [9]:

$$\Delta \equiv \Delta_{p-1} = \left\{ \mathbf{x} \in \mathbb{R}^p : \mathbf{x} > \mathbf{0}, \mathbf{1}_p^\top \mathbf{x} = 1 \right\}. \quad (1)$$

Generative models for HTS-derived compositional data commonly respect compositionality either by transforming the data into Euclidean space through log-ratio or similar transformations [10, 11], or by using distributions directly defined on the probability simplex. The Dirichlet distribution is a popular choice due to its relatively simple structure and interpretability [3, 12–14].

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Statistics in Medicine* published by John Wiley & Sons Ltd.

The assumption of independent features (apart from the compositional effect) is, however, a major limitation of the Dirichlet distribution. To allow for more complex dependency structures, [15] proposed the class of logistic normal distributions, which include the estimation of feature-feature interactions. Several lines of research make successful use of logistic normal models (and extensions thereof) for HTS data (see, e.g., [16, 17]) and address the computational challenges in scaling parameter inference to large-scale datasets [18].

Another challenge in generative HTS data modeling is the presence of zeroes. Since the logistic normal distribution requires the underlying data to be positive due to logarithmic transformations, zero entries in the primary data need to be replaced by positive values [19, 20]. Any such procedure inevitably distorts the measured data compositions, especially for rare features with many zero entries [19], resulting in another source of modeling inaccuracy.

In his seminal work, John Aitchison [21] provided a general class of distributions, the A^{p-1} class, that includes the logistic normal and the Dirichlet distribution as special cases. This class forms the basis for more recent models that extend the A^{p-1} class and do not require zero imputation. For low-dimensional data, [22] and [23] introduced the polynomially tilted pairwise interaction (PPI) model, which has properties similar to the Dirichlet distribution at the boundaries of the simplex. The class of a-b power interaction models, introduced by [24], achieves validity on the simplex boundaries by replacing the logarithm with power transformations. These works further use score matching estimation [25] for computationally efficient parameter inference, reducing the estimation problem to solving a (regularized) quadratic optimization problem. However, both the PPI and the a-b power interaction model currently only allow to model a homogeneous sample population and cannot describe the differences between groups of samples.

A central task in HTS data analysis is the detection of significant differences in the feature composition, given environmental, clinical, or host-specific perturbations or variations. This problem, also known as differential abundance (DA) testing, faces the same challenges as generative modeling [2, 8]. While compositionality and zero handling are discussed in most state-of-the-art DA testing methods [26–28], only a few methods explicitly include interactions between compositional features in their testing procedure [29]. Such interactions, however, may contribute to the false discovery of certain features that are not directly impacted by the perturbations or covariate changes, but simply strongly correlate with the differentially abundant feature. Consider a composition of five microbial taxa a, b, c, d, e , where a and b have a symbiotic relationship and their abundances are highly correlated (Figure 1a). A treatment now targets taxon a and causes a decline in its population. This will, in turn, cause the abundance of taxon b to also decrease, although it was not directly influenced by the treatment. Classical DA testing methods will not be able to discern between these primary and secondary effects caused by the treatment, detecting both a and b as differentially abundant.

In this work, we present a new DA framework, termed `cosmoDA` (**compositional score matching optimization for Differential Abundance analysis**), that addresses the challenge of feature

interactions in DA testing. `cosmoDA` is based on the a-b power interaction models from Reference [24] and introduces a linear covariate effect on the location vector, thus enabling the inclusion of sample group indicators or continuous covariates of interest. We provide a framework for assessing the significance of the estimated covariate effects, which, in the case of group indicator variables, allows principled compositional differential abundance testing. A similar covariate-extended model was introduced for low-dimensional compositional data by [30], albeit only for the special case of the logistic normal model. In the a-b power interaction models, maximum likelihood estimation is not possible due to the intractability of computing the normalizing constant. We thus resort to the score matching framework [25]. By carefully studying the structural properties of the underlying score matching objective, our extended estimation framework retains the favorable quadratic nature of the underlying optimization problem with negligible computational overhead. Regularization on the interaction effects further ensures model identifiability and selection of the most important correlation patterns. The characteristics of the a-b power interaction model thus ensure that feature interactions are adequately considered and zero entries in the underlying data do not need to be replaced or imputed.

The remainder of the paper is structured as follows. In the next section, we introduce `cosmoDA` as an extension of the a-b power interaction model describes the score matching estimation framework and explains how the power transformation makes zero replacement obsolete. We then describe the model regularization framework, sketch the computational implementation, and introduce the differential abundance testing framework `cosmoDA`. Section 3 provides several simulated data benchmarks that showcase the ability of `cosmoDA` to (i) correctly estimate sparse interaction matrices in the presence of covariates and (ii) reduce the false discovery rate in differential abundance testing compared to other state-of-the-art methods. We investigate the impact of different power transformations on differential abundance in real scRNA-seq and 16S rRNA sequencing data in Section 4 and provide a data-driven method to select the power exponents in practice. Section 4 discusses the results, highlights strengths and limitations of the work, and provides guidelines for future research. `cosmoDA` is available as a Python package at <https://github.com/bio-datascience/cosmoDA>.

2 | Methods

We consider to model compositional matrices \mathbf{X} where each row $\mathbf{x}^{(i)} \in \Delta$, $i = 1, \dots, n$, represents a sample, and each column \mathbf{x}_j , $j = 1, \dots, p$, represents the j th compositional feature. We are motivated by the large number of available biological “compositional” count matrices $\tilde{\mathbf{X}} \in \mathbb{N}_0^{n \times p}$ derived from HTS experiments. Important instances include 16S rRNA amplicon sequencing data, where each feature represents read counts associated with a microbial taxon [8] and scRNA-Seq experiments, where each feature represents a certain cell-type proportion, as derived from clustered transcriptional profiles [3, 4]. Due to the compositional nature of the derived count data, a common approach is to scale each observation $\tilde{\mathbf{x}}^{(i)}$ by its library size $S^{(i)} = \sum_{j=1}^p \tilde{x}_j^{(i)}$ to obtain relative abundance samples $\mathbf{x}^{(i)} \in \Delta$ [8].

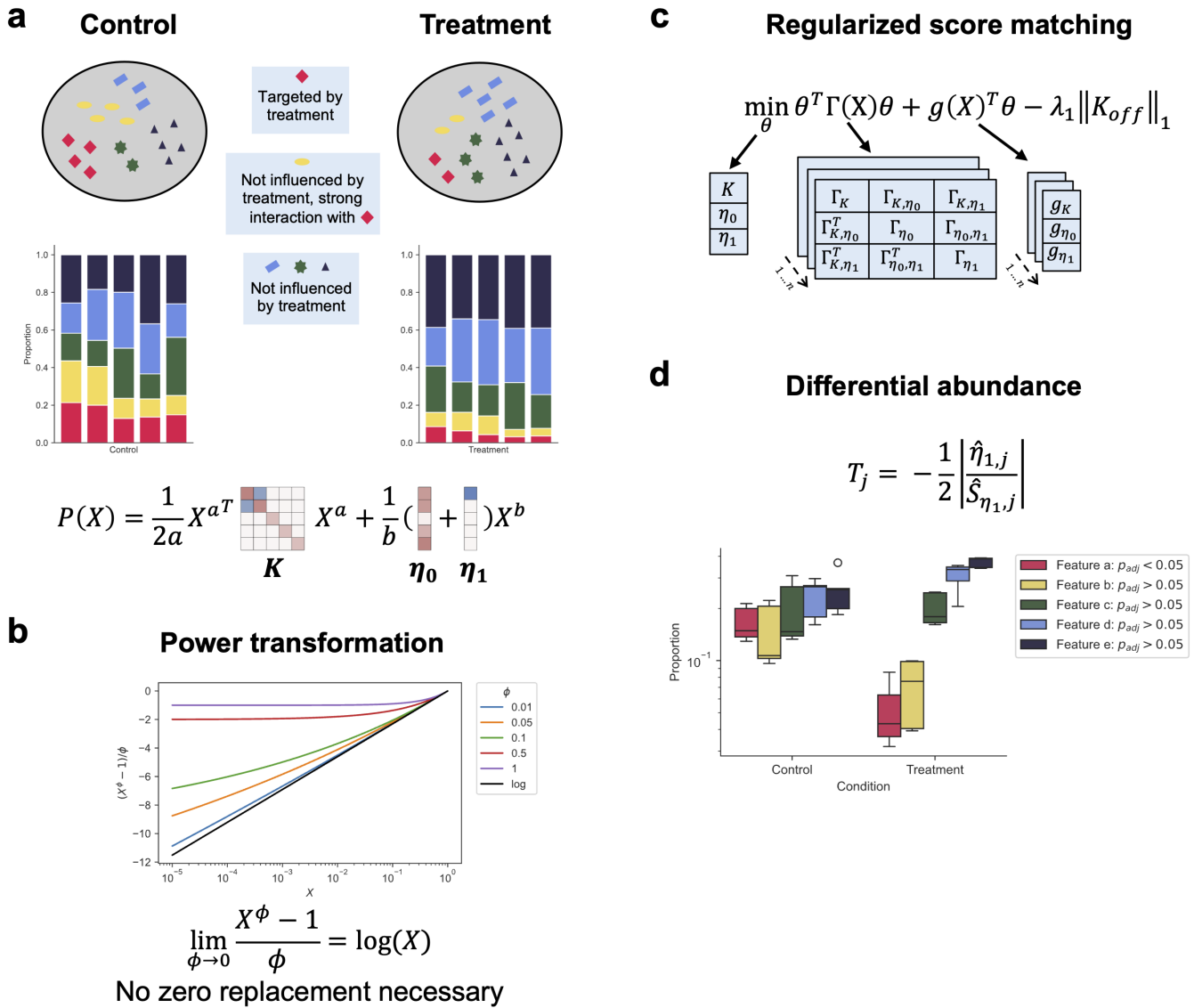


FIGURE 1 | **cosmoDA** allows to perform generative modeling and differential abundance testing on compositional data with feature interactions. (a) Interactions between features can alter the abundance of features although they are not directly associated with the condition. **cosmoDA** is able to accurately distinguish primary from secondary effects by inferring pairwise feature interactions in addition to the effects associated with the condition. (b) Power transformations allow to analyze compositional data without the imputation of zero values. For decreasing exponents, the Box–Cox transformation converges to the logarithm. (c) **cosmoDA** uses regularized score matching for parameter inference. The optimization problem, therefore, reduces to a quadratic function with parameters Γ and g defined by averaging over all samples. (d) Differential abundance testing in **cosmoDA** uses a studentized test statistic. Only the feature primarily associated with the condition (Feature a) has a small adjusted p -value.

2.1 | The Covariate-Extended A-B Power Interaction Model

Following the proposal in Reference [24], we model samples in \mathbf{X} through the *a-b power interaction model* on the $(p-1)$ -dimensional simplex Δ . The unnormalized probability density for one sample $\mathbf{x} \equiv \mathbf{x}^{(i)}$ reads:

$$p_{\eta, \mathbf{K}}(\mathbf{x}) \propto \exp\left(-\frac{1}{2a} \mathbf{x}^{aT} \mathbf{K} \mathbf{x}^a + \frac{1}{b} \boldsymbol{\eta}^T \mathbf{x}^b\right),$$

$$\mathbf{x} \in \Delta; \mathbf{K} = \mathbf{K}^T; \mathbf{K} \mathbf{1}_p = \mathbf{0}_p. \quad (2)$$

Here, interactions between features are modeled through the *interaction matrix* $\mathbf{K} \in \mathbb{R}^{p \times p}$, and the *location vector* $\boldsymbol{\eta} \in \mathbb{R}^p$

describes the base composition of the individual features. This model belongs to the class of exponential family models. Using the conventions in Reference [24], $\mathbf{x}^a \equiv \log(\mathbf{x})$; $1/a \equiv 1$ if $a = 0$, and $\mathbf{x}^b \equiv \log(\mathbf{x})$; $1/b \equiv 1$ if $b = 0$, power interaction models encapsulate several probability distributions as special cases.

With parameter settings $a = b = 0$, the model includes the Dirichlet distribution with the additional constraints $\mathbf{K} = \mathbf{0}$, $\boldsymbol{\eta} > -1$, the logistic normal distribution [15] with the constraints $\mathbf{K} \mathbf{1}_p = \mathbf{0}_p$, $\mathbf{x}^T \mathbf{K} \mathbf{x} > 0 \forall \mathbf{x}$, $\mathbf{1}_p^T \boldsymbol{\eta} = -p$, and Aitchison's A^{p-1} family of distributions [21] with $\mathbf{x}^T \mathbf{K} \mathbf{x} > 0 \forall \mathbf{x}$, $\boldsymbol{\eta} > -1$. For the logistic normal case, the interaction matrix \mathbf{K} is equivalent to the inverse covariance matrix of logratio-transformed data, given specific linear transformations [31]. With parameter settings $a = 1$ and

$b = 0$, the model is equivalent to the PPI distribution [22, 23] (see Appendix A), and with parameter settings $a = b = 1$, the power interaction model is equivalent to the maximum entropy distribution on the simplex with the constraints $\mathbf{K}\mathbf{1}_p = \mathbf{0}_p$, $\mathbf{x}^T \mathbf{K} \mathbf{x} > 0 \forall \mathbf{x}$, as derived in Reference [32].

As stated in Theorem 1 from Reference [24], the probability density in Equation (2) is proper if either

- $a > 0, b > 0$;
- $a > 0, b = 0, \eta_j > -1 \forall j$;
- $a = 0, b = 0, \log(\mathbf{x})^T \mathbf{K} \log(\mathbf{x}) > 0 \forall \mathbf{x} \in \Delta$;
- $a = 0, b > 0, \log(\mathbf{x})^T \mathbf{K} \log(\mathbf{x}) \geq 0 \forall \mathbf{x} \in \Delta$.

We next extend the original proposal of the a-b power interaction model by including a (continuous or binary) covariate vector $\mathbf{y} \in \mathbb{R}^n$ (or $\mathbf{y} \in \{0, 1\}^n$, respectively) in the model. The covariate describes, for example, a concurrently measured quantity of interest for each sample, or, more relevant in our context, a condition-specific indicator vector. We model the influence of $\mathbf{y} \equiv \mathbf{y}^{(i)}$ on $\mathbf{x}^{(i)}$ by introducing a linear model on the location vector $\boldsymbol{\eta}$:

$$\boldsymbol{\eta} = \boldsymbol{\eta}_0 + \mathbf{y}\boldsymbol{\eta}_1. \tag{3}$$

Plugging this model into Equation (2) results in the covariate-extended a-b power interaction model:

$$p_{\boldsymbol{\eta}, \mathbf{K}}(\mathbf{x}) \propto \exp\left(-\frac{1}{2a} \mathbf{x}^a \mathbf{K} \mathbf{x}^a + \frac{1}{b} (\boldsymbol{\eta}_0 + \mathbf{y}\boldsymbol{\eta}_1)^T \mathbf{x}^b\right),$$

$$\mathbf{x} \in \Delta; \mathbf{K}\mathbf{1}_p = \mathbf{K}^T \mathbf{1}_p = \mathbf{0}_p. \tag{4}$$

This formulation of the model assumes that all samples stem from an overall population with fixed interaction matrix \mathbf{K} , but allows the proportions of features, described by $\boldsymbol{\eta}$, to be dependent on the measured covariate \mathbf{y} . For the probability density of the covariate-extended a-b power interaction model to be proper, the same conditions hold as for the model in Equation (2), replacing $\boldsymbol{\eta}$ with $\boldsymbol{\eta}_0 + \mathbf{y}\boldsymbol{\eta}_1$.

The model in Equation (4) forms the basis for our differential abundance testing framework **cosmoDA** (**compositional score matching optimization for Differential Abundance analysis**). In case where \mathbf{y} represents a binary group indicator, for example, case versus control samples, **cosmoDA** fits the data to the model and tests for significant changes of the individual components of $\boldsymbol{\eta}_1$. Figure 1 provides a conceptual overview of **cosmoDA**. Before detailing the specific test statistics, we describe the underlying parameter estimation framework.

2.2 | Model Estimation

2.2.1 | Score Matching for A-B Power Interaction Models

Efficient parameter estimation for the a-b power interaction models Equation (2) through generalized score matching [25, 33] was proposed by [24]. While other estimation approaches (e.g., pseudo-likelihood estimation as used by [32]

in maximum-entropy models) might also provide efficient parameter estimation for our model, [34] and [35] showed score matching to be preferential for similar distributions from the exponential family. Given an (unknown) true data distribution P_0 with density p_0 and a family of distributions of interest $\mathcal{P}(D)$, score matching tries to find $P \in \mathcal{P}(D)$ with density p such that the Hyvärinen divergence between the gradients of the logarithm of the densities of P_0 and P is minimized:

$$\frac{1}{2} \int_D p_0(\mathbf{x}) \left\| \nabla \log p(\mathbf{x}) \odot \tilde{\mathbf{h}}^{1/2}(\mathbf{x}) - \nabla \log p_0(\mathbf{x}) \odot \tilde{\mathbf{h}}^{1/2}(\mathbf{x}) \right\|_2^2 d\mathbf{x}, \tag{5}$$

where $\tilde{\mathbf{h}}(\mathbf{x}) = (\tilde{h}_1(x_1), \dots, \tilde{h}_p(x_p))$ is a weight function. Yu et al. [36] show that score matching can be performed on domains with positive Lebesgue measure in \mathbb{R}^p by setting $\tilde{\mathbf{h}}$ such that $\nabla \log p(\mathbf{x}) \odot \tilde{\mathbf{h}}^{1/2}(\mathbf{x})$ does not vanish at the boundaries of the domain.

Yu et al. [24] adapted the generalized score matching framework for the a-b power interaction models on the $(p - 1)$ -dimensional probability simplex in \mathbb{R}^p by profiling out the last coordinate $x_p \equiv 1 - \sum_{j=1}^{p-1} x_j$, similar to the additive log-ratio transformation. We follow this approach, setting $\tilde{h}_j(\mathbf{x}) = (h_j \circ \varphi_j)(\mathbf{x})$ with $h_j(\mathbf{x}) = x_j^c$ and $\varphi_j(\mathbf{x}) = \min\{x_j, x_p, C_j\}$ and fixing $C_j = 1$ and $c = 2$, as recommended by [24]. This results in a weight function $\tilde{h}_j(\mathbf{x}) = \min\{x_j, x_p\}^2$. With $p(\mathbf{x})$ from the family of a-b power interaction models, the following mild assumptions hold [24]:

1. $p_0(x_j; \mathbf{x}_{-j}) h_j(\varphi_j(\mathbf{x}_j)) \partial_j \log p(x_j; \mathbf{x}_{-j}) \Big|_{x_j \nearrow b_k(\mathbf{x}_{-j})^-} = 0$ for all $k = 1, \dots, K_j(\mathbf{x}_{-j})$ and $\mathbf{x}_{-j} \in S_{-j,D}$ for all $j = 1, \dots, p$;
2. $\int_D p_0(\mathbf{x}) \left\| \nabla \log p(\mathbf{x}) \odot (\mathbf{h} \circ \boldsymbol{\varphi})^{1/2}(\mathbf{x}) \right\|_2^2 d\mathbf{x} < +\infty$,
 $\int_D p_0(\mathbf{x}) \left\| [\nabla \log p(\mathbf{x}) \odot (\mathbf{h} \circ \boldsymbol{\varphi})(\mathbf{x})] \right\|_1 d\mathbf{x} < +\infty$.
3. $\forall j = 1, \dots, p$ and almost everywhere $\mathbf{x}_{-j} \in S_{-j,D}$, the component function h_j of \mathbf{h} is absolutely continuous in any bounded sub-interval of the section $C_{j,D}(\mathbf{x}_{-j})$.

Therefore, a consistent estimator of the loss function Equation (5) follows as a sample- and feature-wise sum over the entire dataset:

$$\hat{L}_h(P) = \frac{1}{2} \sum_{j=1}^p \sum_{i=1}^n \frac{1}{2} (h_j \circ \varphi_j)(\mathbf{X}^{(i)}) \cdot [\partial_j \log p(\mathbf{X}^{(i)})]^2 + \partial_j [(h_j \circ \varphi_j)(\mathbf{X}^{(i)}) \cdot \partial_j \log p(\mathbf{X}^{(i)})], \tag{6}$$

where $\mathbf{X}^{(i)}$, $1 \leq i \leq n$, form an i.i.d. sample from the unknown data distribution P_0 and P is an a-b power interaction model with unnormalized density as described in Equation (2). Aggregating \mathbf{K} and $\boldsymbol{\eta}$ to $\boldsymbol{\theta} = (\text{vec}(\mathbf{K}), \boldsymbol{\eta})$ and defining P_θ and its density p_θ accordingly shows that the power interaction model without covariate Equation (2) follows an exponential-family-type model

$$\log p_\theta(\mathbf{x}) = \boldsymbol{\theta}^T \mathbf{t}(\mathbf{x}) - \psi(\boldsymbol{\theta}) + b(\mathbf{x}), \mathbf{x} \in \Delta, \tag{7}$$

where the function $\mathbf{t}(\cdot)$ denotes the function for the sufficient statistics, $\psi(\cdot)$ the cumulant function, and $b(\cdot)$ the logarithm of the base measure, respectively.

Then, \hat{L}_h can be reformulated as a quadratic optimization problem:

$$\hat{L}_h(P_\theta) = \frac{1}{2} \theta^\top \Gamma(\mathbf{x}) \theta - \mathbf{g}(\mathbf{x})^\top \theta + \text{const.} \quad (8)$$

with $\Gamma(\mathbf{x}) \in \mathbb{R}^{r \times r}$ and $\mathbf{g}(\mathbf{x}) \in \mathbb{R}^r$ are sample averages of known functions in \mathbf{x} only.

Analogously, the same considerations hold for the covariate-extended a-b power interaction model Equation (4), substituting $\boldsymbol{\eta}$ with $\boldsymbol{\eta}_0 + y\boldsymbol{\eta}_1$. This substitution does not yet provide individual estimates of $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_0$, though, which are required for differential abundance testing. To obtain these individual estimates, a look at the exact derivation of Γ and \mathbf{g} , as described by [24], is necessary. We first split the location vector into its two parts $\boldsymbol{\eta}_0$ and $\boldsymbol{\eta}_1$, and set $\boldsymbol{\theta} = (\text{vec}(\mathbf{K}), \boldsymbol{\eta}_0, \boldsymbol{\eta}_1)$. After dropping the last coordinate by assuming $x_p \equiv 1 - \mathbf{1}_{p-1}^\top \mathbf{x}_{-p}$ as above, the first and second partial derivatives for the covariate-less model (Equations (4.1) and (4.2) in Reference [24]) can easily be adapted to the covariate-extended model:

$$\begin{aligned} \partial_j \log p(\mathbf{x}_{-p}) &= -(\boldsymbol{\kappa}_{\cdot j}^\top \mathbf{x}^a) x_j^{a-1} + (\boldsymbol{\kappa}_{\cdot p}^\top \mathbf{x}^a) x_p^{a-1} + \eta_j x_j^{b-1} - \eta_p x_p^{b-1}, \\ &= -(\boldsymbol{\kappa}_{\cdot j}^\top \mathbf{x}^a) x_j^{a-1} + (\boldsymbol{\kappa}_{\cdot p}^\top \mathbf{x}^a) x_p^{a-1} \\ &\quad + \eta_{0,j} x_j^{b-1} - \eta_{0,p} x_p^{b-1} + y_j \eta_{1,j} x_j^{b-1} - y_p \eta_{1,p} x_p^{b-1} \quad (9) \end{aligned}$$

$$\begin{aligned} \partial_{jj} \log p(\mathbf{x}_{-p}) &= -(a-1) \left[(\boldsymbol{\kappa}_{\cdot j}^\top \mathbf{x}^a) x_j^{a-2} + (\boldsymbol{\kappa}_{\cdot p}^\top \mathbf{x}^a) x_p^{a-2} \right] \\ &\quad - a \left[\kappa_{jj} x_j^{2a-2} + \kappa_{pp} x_p^{2a-2} + 2\kappa_{jp} x_j^{a-1} x_p^{a-1} \right] \\ &\quad + (b-1) \left[\eta_j x_j^{b-2} + \eta_p x_p^{b-2} \right] \\ &= -(a-1) \left[(\boldsymbol{\kappa}_{\cdot j}^\top \mathbf{x}^a) x_j^{a-2} + (\boldsymbol{\kappa}_{\cdot p}^\top \mathbf{x}^a) x_p^{a-2} \right] \\ &\quad - a \left[\kappa_{jj} x_j^{2a-2} + \kappa_{pp} x_p^{2a-2} + 2\kappa_{jp} x_j^{a-1} x_p^{a-1} \right] \\ &\quad + (b-1) \left[\eta_{0,j} x_j^{b-2} + \eta_{0,p} x_p^{b-2} \right] \\ &\quad + (b-1) \left[y_j \eta_{1,j} x_j^{b-2} + y_p \eta_{1,p} x_p^{b-2} \right] \quad (10) \end{aligned}$$

Plugging these definitions into the loss function in Equation (6) and rearranging the individual terms in the same way as in Reference [24] yields Γ and \mathbf{g} as follows (Figure 1c):

$$\begin{aligned} \Gamma &\equiv \begin{bmatrix} \Gamma_{\mathbf{K}} & \Gamma_{\mathbf{K},\boldsymbol{\eta}_0} & \Gamma_{\mathbf{K},\boldsymbol{\eta}_1} \\ \Gamma_{\mathbf{K},\boldsymbol{\eta}_0}^\top & \Gamma_{\boldsymbol{\eta}_0} & \Gamma_{\boldsymbol{\eta}_0,\boldsymbol{\eta}_1} \\ \Gamma_{\mathbf{K},\boldsymbol{\eta}_1}^\top & \Gamma_{\boldsymbol{\eta}_0,\boldsymbol{\eta}_1}^\top & \Gamma_{\boldsymbol{\eta}_1} \end{bmatrix} \in \mathbb{R}^{(p^2+2p) \times (p^2+2p)}, \\ \mathbf{g} &\equiv (\text{vec}(\mathbf{g}_{\mathbf{K}}), \mathbf{g}_{\boldsymbol{\eta}_0}, \mathbf{g}_{\boldsymbol{\eta}_1}) \in \mathbb{R}^{p^2+2p}, \quad (11) \end{aligned}$$

where Γ and \mathbf{g} have a block structure with $\Gamma_{\mathbf{K}} \in \mathbb{R}^{p^2 \times p^2}$, $\Gamma_{\mathbf{K},\boldsymbol{\eta}_0} \in \mathbb{R}^{p^2 \times p}$, $\Gamma_{\mathbf{K},\boldsymbol{\eta}_1} \in \mathbb{R}^{p^2 \times p}$, $\Gamma_{\boldsymbol{\eta}_0} \in \mathbb{R}^{p \times p}$, $\Gamma_{\boldsymbol{\eta}_0,\boldsymbol{\eta}_1} \in \mathbb{R}^{p \times p}$, $\Gamma_{\boldsymbol{\eta}_1} \in \mathbb{R}^{p \times p}$, and $\mathbf{g}_{\mathbf{K}} \in \mathbb{R}^{p^2}$, $\mathbf{g}_{\boldsymbol{\eta}_0} \in \mathbb{R}^p$, $\mathbf{g}_{\boldsymbol{\eta}_1} \in \mathbb{R}^p$.

The exact derivations are shown in Appendix B. By recognizing that each entry of Γ and \mathbf{g} can be written as a mean over all samples, the elements related to $\boldsymbol{\eta}_1$ can be computed directly from the elements related to $\boldsymbol{\eta}_0$:

$$\begin{aligned} \Gamma_{\mathbf{K},\boldsymbol{\eta}_1} &= \frac{1}{n} \sum_{i=1}^n y \Gamma_{\mathbf{K},\boldsymbol{\eta}_0}^{(i)} \\ \Gamma_{\boldsymbol{\eta}_0,\boldsymbol{\eta}_1} &= \frac{1}{n} \sum_{i=1}^n y \Gamma_{\boldsymbol{\eta}_0}^{(i)} \end{aligned}$$

$$\begin{aligned} \Gamma_{\boldsymbol{\eta}_1} &= \frac{1}{n} \sum_{i=1}^n y^2 \Gamma_{\boldsymbol{\eta}_0}^{(i)} \\ \mathbf{g}_{\boldsymbol{\eta}_1} &= \frac{1}{n} \sum_{i=1}^n y \mathbf{g}_{\boldsymbol{\eta}_0}^{(i)}. \end{aligned}$$

Therefore, the computational overhead for computing the additional sub-matrices and sub-vectors related to $\boldsymbol{\eta}_1$ is negligible. Still, the addition of p dimensions to the optimization problem Equation (8) increases the problem dimensionality from $p^2 + p$ to $p^2 + 2p$ compared to the covariate-less model.

2.2.2 | Model Identifiability Through Regularization

Since the number of parameters in the power interaction model scales quadratically with p , real HTS data applications are in the high-dimensional regime with more parameters than samples, that is, $p^2 + 2p > n$. To enable model identification, we place a ℓ_1 regularization penalty on the off-diagonal elements \mathbf{K}_{off} of \mathbf{K} :

$$\hat{L}_{h,C,\lambda_1,\delta}(P_\theta) = \frac{1}{2} \theta^\top \Gamma_\delta(\mathbf{x}) \theta - \mathbf{g}(\mathbf{x})^\top \theta + \lambda_1 \|\text{vec}(\mathbf{K}_{\text{off}})\|_1. \quad (12)$$

As defined in Section 2.2.1, $\boldsymbol{\theta} = (\text{vec}(\mathbf{K}), \boldsymbol{\eta}_0, \boldsymbol{\eta}_1)$ comprises all model parameters, and P_θ denotes the power interaction model with probability density p_θ defined in Equation (4). Following [24], we multiply the diagonal entries of $\Gamma(\mathbf{x})$ corresponding to \mathbf{K} by a factor $\delta > 1$ to avoid an unbounded loss function. We denote $\Gamma(\mathbf{x})$ with scaled diagonal entries as $\Gamma_\delta(\mathbf{x})$. Here, we use the default value from the implementation of [24], $\delta = 2 - \frac{1}{1+4e \max(6 \log(p)/n, \sqrt{6 \log(p)/n})}$. In cases where $p \gg n$, the entries of $\boldsymbol{\eta}_0$ and $\boldsymbol{\eta}_1$ can be penalized as well with a regularization parameter λ_2 :

$$\begin{aligned} \hat{L}_{h,C,\lambda_1,\lambda_2,\delta}(P_\theta) &= \frac{1}{2} \theta^\top \Gamma_\delta(\mathbf{x}) \theta - \mathbf{g}(\mathbf{x})^\top \theta + \lambda_1 \|\text{vec}(\mathbf{K}_{\text{off}})\|_1 \\ &\quad + \lambda_2 \|\boldsymbol{\eta}_0\|_1 + \lambda_2 \|\boldsymbol{\eta}_1\|_1 \quad (13) \end{aligned}$$

Furthermore, assuming \mathbf{K} to be sparse matches the widely popular view of sparse association networks between microbial features or cell types (see, e.g., [37]). In the following, we will focus our attention on models without regularization on the location parameter.

Theorem 4 in Reference [24] details that for a proper a-b power interaction model (see Section 2.1), a boundary function $\tilde{h}_j(\mathbf{x})$ as described in Section 2.2.1, and under some mild conditions to the regularization strength λ and diagonal multiplier δ , the estimation error is bounded closely to the true values of \mathbf{K} and $\boldsymbol{\eta}$. As our extension only divides $\boldsymbol{\eta}$ into two components $\boldsymbol{\eta}_0$ and $\boldsymbol{\eta}_1$, similar guarantees hold for the covariate-extended a-b power interaction model.

2.2.3 | Computational Implementation

The regularized score matching loss $\hat{L}_{h,C,\lambda,\delta}(P_\theta)$ Equation (12) represents a (large-scale) ℓ_1 -penalized quadratic optimization problem that can be numerically solved with a variety of optimization methods. Here, we follow [24] and employ a proximal coordinate descent scheme (see also Algorithm 2 in Reference [35]). This algorithm also covers the covariate-extended a-b

ALGORITHM 1

Input: Initial estimate $\hat{\theta}^{(0)}$
Input: t_{max} , maximum number of iterations
Input: ϵ , the maximal tolerance level

- 1: Initialize $t \leftarrow 1$
- 2: Initialize $C \leftarrow \epsilon + 1$ (C stands for convergence criteria)
- 3: **while** $C > \epsilon$ or $t < t_{max}$ **do**
- 4: $\hat{\theta}^{(t)} \leftarrow \hat{\theta}^{(t-1)}$
- 5: **for** $j \leftarrow 1, 2, \dots, s$ **do**
- 6: $\hat{\theta}_j^{(t)} \leftarrow \text{Soft} \left(\frac{-(\Gamma_{\delta}(\mathbf{x})_{-j})^T \hat{\theta}_j^{(t-1)} - g(\mathbf{x})_j}{\Gamma_{\delta}(\mathbf{x})_{jj}}, \frac{\lambda}{\Gamma_{\delta}(\mathbf{x})_{jj}} \right)$.
- 7: **end for**
- 8: $C \leftarrow \|\hat{\theta}^{(t)} - \hat{\theta}^{(t-1)}\|_1$
- 9: $t \leftarrow t + 1$
- 10: **end while**

power interaction model and is described in Algorithm 1. Here, s is the dimensionality of θ and $\text{Soft}(\cdot)$ is the softmax function. The default settings in `cosmoDA` are $\epsilon = 10^{-1}$ and $t_{max} = 1000$.

At its core, our Python implementation of Algorithm 1 uses the C implementation from the `genscore` package [24, 34] and included in the `cosmoDA` Python package. The `cosmoDA` package also provides an interface for a-b power interaction models that is equivalent to the R interface in the `genscore` package.

2.3 | Differential Abundance Testing

One of the key objectives of `cosmoDA` is to determine the statistical significance of the covariate effects $\eta_{1,j}$ for every feature $j = 1 \dots p$. Here, we combine results from References [27] and [22] to test the null hypothesis

$$H_0 : \eta_{1,j} = 0 \quad \text{against the alternative} \quad H_1 : \eta_{1,j} \neq 0.$$

Let $\hat{\theta} = (\text{vec}(\hat{\mathbf{K}}), \hat{\eta}_0, \hat{\eta}_1)$ be the parameter estimates obtained from the score matching estimation framework. Sealy and Wood [22] show that, under certain technical conditions and assumptions (see Theorems 1 and 2 in Reference [22]) the quantity $\hat{\mathbf{S}} = \Gamma^{-1}(\mathbf{x}) \hat{\Sigma}_0 \Gamma^{-1}(\mathbf{x})$ yields a consistent estimator for $\text{Var}(\hat{\theta})$. In `cosmoDA`, we estimate $\hat{\Sigma}_0$ as follows:

$$\hat{\Sigma}_0 = \frac{1}{n} \sum_{i=1}^n (\tilde{\Gamma}_{\delta}^{(i)}(\mathbf{x}) \hat{\theta} - \tilde{\mathbf{g}}^{(i)}(\mathbf{x})) (\tilde{\Gamma}_{\delta}^{(i)}(\mathbf{x}) \hat{\theta} - \tilde{\mathbf{g}}^{(i)}(\mathbf{x}))^T, \quad (14)$$

where $\tilde{\Gamma}_{\delta}^{(i)}(\mathbf{x})$ and $\tilde{\mathbf{g}}^{(i)}(\mathbf{x})$ are the components of $\tilde{\Gamma}_{\delta}(\mathbf{x})$ and $\mathbf{g}(\mathbf{x})$ corresponding to the i -th sample. By selecting the components of $\hat{\mathbf{S}}$ corresponding to $\eta_{1,j}$, we derive the studentized test statistic

$$T_j = \hat{\eta}_{1,j} / \hat{S}_{\eta_{1,j}}, \quad (15)$$

which approximately follows a t -distribution with $n - 3$ degrees of freedom. The corresponding asymptotic p -values read:

$$p_j = 2F_{t,n-3}(-|T_j|), \quad (16)$$

where $F_{t,n-3}$ is the cumulative distribution function of the t -distribution with $n - 3$ degrees of freedom [27]. In `cosmoDA`,

raw p -values are adjusted for multiple testing using the Benjamini–Hochberg procedure [38] (see Figure 1d for illustration).

We empirically verified the distribution of test statistics in Figure F1 with the help of data from our model comparison benchmark with $p = 11$ (Section 3.2). There, we plotted the distribution of test statistics for features following the null hypothesis (i.e., without differential abundance) for both simulated sample sizes $n = 100$ and $n = 1000$. These distributions closely match the densities of the respective t_{n-3} -distributions.

2.4 | Model Selection and Hyperparameter Tuning

To make `cosmoDA` fully data-adaptive, we provide several strategies to select the hyperparameters of the framework. We first describe regularization parameter selection, followed by a novel data-driven approach to select the exponents a and b in a-b power interaction models.

2.4.1 | Regularization Parameter Selection

`cosmoDA` provides several model selection methods to determine the regularization parameter λ_1 (and λ_2 , respectively). The default strategy is k -fold cross-validation ($k = 5$) with the 1SE rule [39]. Here, the largest λ_1 value is chosen that lies within one standard error band of the λ_1 that minimizes the cross-validated regularized score matching loss Equation (12). The range of the λ_1 -path is chosen to cover the whole range of possible sparsity of \mathbf{K} , that is, from a fully dense \mathbf{K} to a diagonal \mathbf{K} (see Figure F16b). This range depends on the dimensions of the dataset at hand and the chosen power transformation (see Appendix C). Per default, our implementation considers 100 λ_1 -values log-linearly spaced in the interval $[10^{-6}, 1]$.

`cosmoDA` also allows λ_1 to be selected via the extended Bayesian Information Criterion (eBIC, [40]). Following [34], the eBIC for the a-b power interaction model reads:

$$\begin{aligned} \text{eBIC}_{\gamma}(\text{vec}(\mathbf{K}_{\text{off}})) &= S(\text{vec}(\mathbf{K}_{\text{off}})) \log(n) - 2 \log(\hat{L}_{h,C,\lambda_1,\delta}(P_{\theta})) \\ &+ 2\gamma \|\text{vec}(\mathbf{K}_{\text{off}})\|_1, \end{aligned} \quad (17)$$

where $S(\text{vec}(\mathbf{K}_{\text{off}}))$ denotes the size of the support of $\text{vec}(\mathbf{K}_{\text{off}})$. The default γ value is $\gamma = 0.5$.

2.4.2 | Data-Driven Selection of A-B Powers

A key strength of a-b power interaction models is their seamless applicability to compositional data with excess zeros. While the limiting case $a = b = 0$ (i.e., log-transforming the data) requires a strategy for zero replacement or zero imputation [19, 20] with potentially detrimental effects for downstream analysis [41], we propose a data-driven tuning strategy for power interaction models with powers $a > 0$ and $b > 0$ that keeps the original data unaltered. For simplicity, we consider the setting $a = b$.

We first note that the power transformations in the models (2) and (4) are similar to the Box–Cox transformation [42] of the form:

$$x_\phi = \begin{cases} \frac{1}{\phi}(x^\phi - 1), & \text{if } \phi > 0 \\ \log x, & \text{if } \phi = 0, \end{cases} \quad (18)$$

with $\lim_{\phi \rightarrow 0} \frac{1}{\phi}(x^\phi - 1) = \log(x)$ (see also Figure 1b for illustration). The Box–Cox transformation and the power transformation used in a-b power interaction models Equations (2), (4) are, however, not equivalent due to the -1 term in the Box–Cox transformation. By introducing scaling factors for the score matching elements in Equation (11), we can nevertheless achieve the same asymptotic approximation to the logarithm as the Box–Cox transformation (see Appendix C for details).

While ϕ is typically tuned to make the transformed data approach normality, we follow a geometric strategy inspired by the one presented in Reference [43]. Specifically, we determine $\phi = a = b$ to let the resulting “geometry” of transformed data be as similar as possible to the appropriate log-ratio geometry. This is achieved by maximally aligning the principal component (PC) embedding of log-ratio transformed data with imputed zeros and the PC embedding of a power transformation with parameter ϕ of the data with zero entries [44]. Maximal alignment is defined as the highest Procrustes correlation of both embeddings over a range of values $\phi \in]0, 1[$. In the case of power interaction models, it is natural to select a power that closely matches the geometry of data after the additive log-ratio (ALR) transform, since a-b power interaction models with $\phi = a = b = 0$ are a generalization of Aitchison’s A^{p-1} distributions after ALR transformation of the data [15]. Since equal dimensionality of the ALR and power-transformed data is required, we append the column $\log(\frac{x_i}{x_p}) = \mathbf{0}_p$ to the ALR transformation of \mathbf{X} before performing PC analysis.

The original procedure to obtain maximal Procrustes correlation is outlined in Reference [43]. We use the same procedure, but with different input matrices. Let \mathbf{X}_ϕ be the Box–Cox-like transformed data with

$$\mathbf{X}_{\phi,j} = \frac{1}{\phi} \left(p \frac{\mathbf{X}_j^\phi}{\sum_{k=1}^p \mathbf{X}_k^\phi} - 1 \right), \quad (19)$$

and \mathbf{X}_{ALR} is the ALR-transformed data (with pseudocount 0.5 for all zeros) with column $\mathbf{0}_p$ appended.

We compute the Procrustes correlation r_ϕ between the two data matrices as follows:

- (i) Matrix normalization: $\mathbf{X}_\phi^* = \mathbf{X}_\phi / \sqrt{\text{trace}(\mathbf{X}_\phi^T \mathbf{X}_\phi)}$
 $\mathbf{X}_{ALR}^* = \mathbf{X}_{ALR} / \sqrt{\text{trace}(\mathbf{X}_{ALR}^T \mathbf{X}_{ALR})}$
- (ii) SVD of cross product: $\mathbf{S} = \mathbf{X}_\phi^{*T} \mathbf{X}_{ALR}^* = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$
- (iii) Optimal rotation matrix: $\mathbf{Q} = \mathbf{V} \mathbf{U}^T$
- (iv) Sum of squared errors: $E_\phi = \text{trace}((\mathbf{X}_\phi^* - \mathbf{X}_{ALR}^* \mathbf{Q})^T (\mathbf{X}_\phi^* - \mathbf{X}_{ALR}^* \mathbf{Q}))$
- (v) Procrustes correlation: $r_\phi = \sqrt{1 - E_\phi}$

For a given dataset, the optimal power ϕ^* is determined by $\phi^* := \arg \max_{\phi} r_\phi$ for $\phi \in]0, 1[$.

3 | Simulation Benchmarks

We next provide two simulation studies that benchmark two key features of COSMO DA: (i) sparse recovery of feature interactions in the covariate-extended a-b power interaction model and (ii) identification of differentially abundant features in the presence of feature correlations. The first benchmark complements the extensive covariate-free simulation benchmarks of [24], the second one provides a new realistic semi-synthetic simulation and evaluation setup, incorporating scRNA-Seq data [45].

3.1 | Sparse Recovery of Feature Interactions in the Presence of a Covariate

One of the core strengths of a-b power interaction models is their ability to recover (potentially) sparse feature interaction matrices \mathbf{K} . Yu et al. [24] provides an extensive simulation framework that evaluates the influence of hyperparameters, sample size, and interaction topologies on the recovery performance of the a-b power interaction model. We focus here on evaluating the influence of covariate inclusion on the model’s ability to identify sparse feature interactions. Specifically, we expect interaction recovery to be *independent of covariate inclusion*.

Following [24], we generated compositional data $\mathbf{X} \in \Delta_{p-1}^n$ from an A^{p-1} model with $p = 100$ features using the model in Equation (4) with the constraint that $\mathbf{x}^T \mathbf{K} \mathbf{x} > 0 \forall \mathbf{x}, \boldsymbol{\eta} > -1$. To probe sample size dependencies, we used two scenarios: $n = 80$ and $n = 1000$, respectively. We set $\boldsymbol{\eta}_0 = -\mathbf{1}_p$, and considered banded interaction matrices \mathbf{K} with bandwidths $s = 2$ if $n = 80$ and $s = 7$ if $n = 1000$, as suggested by [24]. We further defined the nonzero off-diagonal entries of \mathbf{K} as $\mathbf{K}_{i,j} = |i - j| / (s + 1) - 1$ for all $i \neq j, 1 \leq |i - j| \leq s$, and the diagonal entries as the negative sum of the off-diagonals, to ensure the sum-to-zero constraint on the rows of \mathbf{K} (Figure F2). This definition slightly deviates from the definition in Reference [24], as the sign of all entries in \mathbf{K} is flipped, but ensures positive definiteness of \mathbf{K} . This modification allows the efficient use of the adaptive rejection sampler for data generation, as provided in the *genscore* R package [34]. For both sample sizes, we generated $R = 50$ replicates of the data.

We applied three different methods for regularized estimation of the underlying interaction matrix \mathbf{K} to all datasets:

- i. The a-b power interaction model ($a = b = 0$) without covariate Equation (2). This model allows the estimation of \mathbf{K} and $\boldsymbol{\eta}_0$. We estimated these models through the implementation in COSMO DA.
- ii. The covariate-extended a-b power interaction model ($a = b = 0$) Equation (4). Here, we used a misspecified \mathbf{y} where each entry is drawn uniformly at random from $\{0, 1\}$. The model allows the estimation of \mathbf{K} , $\boldsymbol{\eta}_0$, and $\boldsymbol{\eta}_1$. We used the implementation in COSMO DA.
- iii. The graphical lasso model on CLR-transformed data, as introduced in SPIEC-EASI [37]. The non-zero entries of

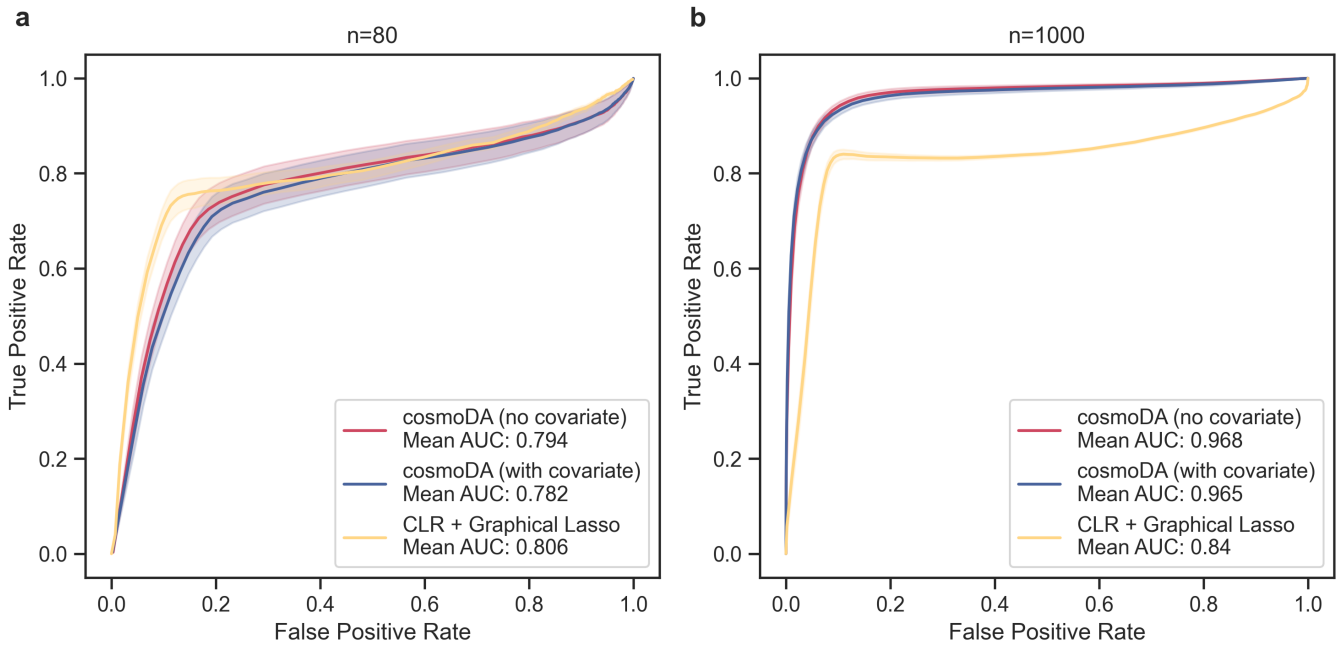


FIGURE 2 | **Recovery of \mathbf{K} improves with sample size and is not impacted by covariate inclusion.** Receiver operating curves for `cosmoDA` with and without covariate effect estimation, as well as CLR transform and graphical lasso for with (a) $n = 80$ and (b) $n = 1000$. The solid lines depict the mean ROC over all 50 generated datasets, and the shaded areas show the standard error.

the resulting sparse inverse covariance matrix serve as a (mis-specified) proxy for \mathbf{K} . We used the implementation from the `glasso` package [46]. Model selection was performed with the extended BIC (eBIC) criterion [40] with $\gamma = 0.25$.

For all three models, we used $n_\lambda = 100$ values for the regularization parameter, log-spaced in the range $10^{-6} < \lambda_1 < 1$, and $k = 5$ cross-validation folds. All score matching estimation parameters were set to the defaults recommended by [24] (see also Section 2.2.3).

To measure recovery performance, we compared the support of the off-diagonal elements of the estimated $\hat{\mathbf{K}}$ and the ground truth \mathbf{K} by calculating the true positive rate (TPR) and true negative rate (TNR), and assessing them through the receiver operating characteristic (ROC) curves.

Figure 2 summarizes the average ROC curves for the two different sample sizes. For $n = 80$ (Figure 2a), we observed that both a-b power interaction models showed almost equivalent ability to reconstruct the interaction matrix (mean AUC 0.782 vs. 0.794). Their performance was slightly worse than the graphical lasso (mean AUC 0.806), especially for false positive rates smaller than 0.2. When increasing the sample size to $n = 1000$, all three methods showed improvements in recovery performance, improving the mean average AUC as well as reducing the variance in results (Figure 2b). As expected, including a covariate in the a-b power interaction model had only a marginal impact on the mean AUC (0.965 vs. 0.968). Contrary to the low sample size case, both a-b power interaction models significantly outperformed the (misspecified) graphical lasso (mean AUC 0.84) across the entire range of regularization strengths.

3.2 | Differential Abundance Testing in the Presence of Correlated Features

To test the effectiveness of `cosmoDA` in detecting differentially abundant features in the presence of realistic feature interactions, we designed the following semi-synthetic simulation benchmark.

We considered a scRNA-seq data set from Reference [45] that derived relative abundance values of $p = 11$ types of peripheral blood mononuclear cells (PBMCs) from overall $n = 352$ samples. The samples come from 260 unique subjects, 162 of whom are patients with systemic lupus erythematosus (SLE) (208 samples) and 98 healthy controls (144 samples). We used these data to estimate realistic base values for the interaction matrix \mathbf{K} and the location vector $\boldsymbol{\eta}$, respectively. The base model is the a-b power interaction model without covariate Equation (2). We set $a = b = 0$ and used $\lambda_1 = 0.043$ as sparsity parameter. We considered the NK cell type as the p th reference component for all power interaction models due to their high abundance and low variance between groups. The resulting interaction matrix \mathbf{K}_B and location vector $\boldsymbol{\eta}_{0,B}$ are shown in Figure 3.

To generate ground-truth differentially abundant cell types, we defined the effect vector $\boldsymbol{\eta}_{1,B} = \tau \mathbf{i} \boldsymbol{\eta}_{0,B}$, where \mathbf{i} is a p -dimensional binary vector that indicates the cell types that are influenced by the condition (i.e., are differentially abundant), and $\tau = (-0.5, -0.3, 0.3, 0.5, 1)$ controls the relative effect size.

Using this model, we considered three differential abundance scenarios: (i) Estimation when the effect is on a rare cell type with (pDC), (ii) effect estimation on an abundant cell type (T4), and (iii) effects on both cell types (pDC and T4). Two different sample sizes ($n = 100$ and $n = 1000$) were considered for each case. For each of the resulting 30 scenarios, we generated

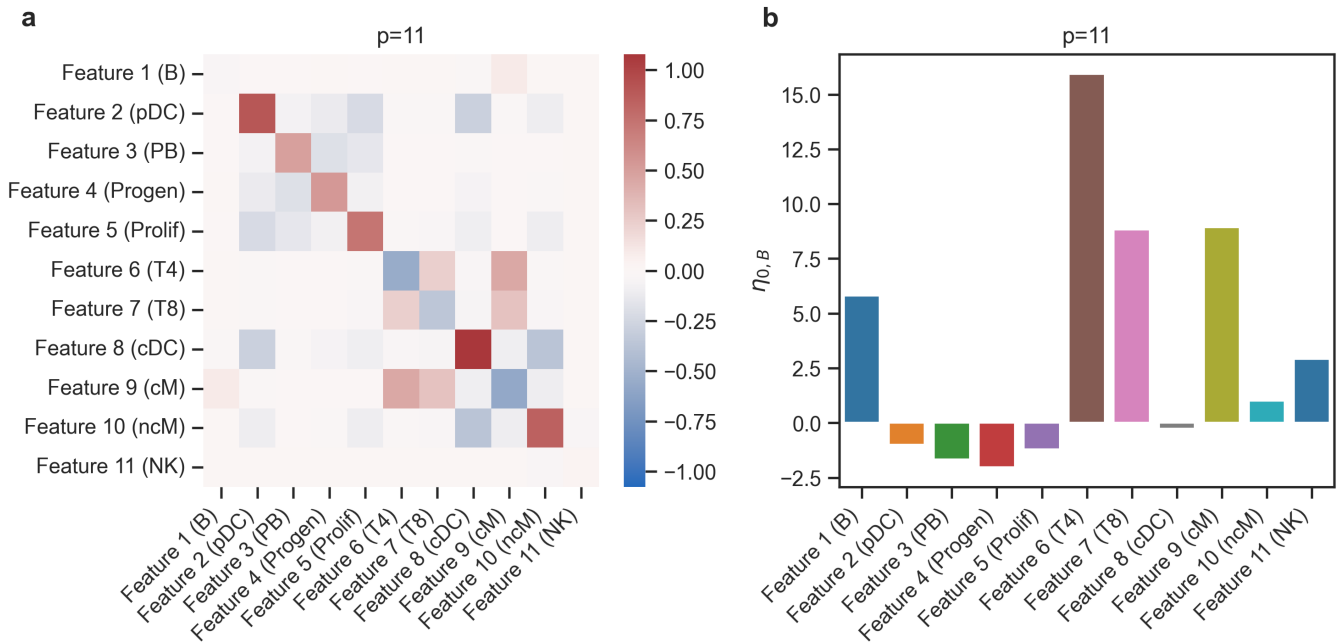


FIGURE 3 | Data generation parameters used for the Differential abundance testing benchmark, $p = 11$. Parameters were generated by running the power interaction model without covariate on the dataset from [45]. The names of the cell types from the original dataset are shown in brackets. (a) Interaction matrix (\mathbf{K}_B), (b) Location vector ($\eta_{0,B}$).

five datasets with $n/2$ control samples ($\mathbf{K} = \mathbf{K}_B$, $\boldsymbol{\eta} = \boldsymbol{\eta}_{0,B}$) and $n/2$ case samples ($\mathbf{K} = \mathbf{K}_B$, $\boldsymbol{\eta} = \boldsymbol{\eta}_{0,B} + \boldsymbol{\eta}_{1,B}$). To simulate these semi-synthetic data sets, we used the adaptive rejection sampler from the *genscore* R package [34]. For completeness, we also considered an identical simulation scenario without correlated features with $\mathbf{K} = \mathbf{0}$. We show and discuss the results of this baseline scenario in Appendix D.

To illustrate the performance of *cosmoDA* for higher-dimensional datasets, we conducted another set of simulations with $p = 99$ features. We constructed the corresponding interaction matrix as a block-diagonal matrix, using the original \mathbf{K}_B matrix in each of the nine blocks (see Figure F4a). Likewise, we stacked the scenario-specific location vectors $\boldsymbol{\eta}_{0,B}$ and $\boldsymbol{\eta}_{1,B}$ nine times to obtain the high-dimensional location vectors (Figure F4b).

We compared the ability of four different DA testing methods to recover differentially abundant features at an expected FDR level of $\alpha = 0.05$:

- DA testing with *cosmoDA* ($a = b = 0$). We used $n_\lambda = 20$ values between 10^{-3} and 2 for λ_1 and 5-fold cross validation with the 1SE rule for model selection. All other parameters were set to default values described in Section 2.2.3.
- ANCOM-BC [26] with default parameters as an example for a common DA testing method without feature interactions. Since ANCOM-BC assumes count data instead of relative abundances, we scaled the simulated data by the median sequencing depth over all samples in the original dataset and rounded to the nearest integer to obtain comparable counts.
- A Dirichlet regression model and subsequent significance test on the regression coefficients, as implemented by [47].

This model serves as a simple baseline that does not respect feature interactions.

- CompDA [29], a recent DA testing method for compositional data, respecting feature interactions via conditional dependency modeling.

Figure 4 summarizes the results for the simulation scenario with the original number of features ($p = 11$). Here, *cosmoDA* showed the overall best ability to recover the true effects (Matthews' correlation coefficient, Figure 4a), especially when the sample size was larger and for the more abundant cell type T4 (see Figure F5). Importantly, *cosmoDA* showed the lowest FDR value in all scenarios (Figure 4b). Although *cosmoDA* was not able to control the FDR at the expected level of 0.05 in every scenario, the methods without consideration of interactions (ANCOM-BC and Dirichlet) detected more false positive features, with FDR levels averaging between 0.2 and 0.7 in most scenarios. Surprisingly, CompDA did not achieve lower FDR values than ANCOM-BC and performed worse than Dirichlet regression in all cases. We observed slightly elevated FDR levels of *cosmoDA* in cases where the DA cell types were not detected, resulting in FDR values of 1 where one feature was falsely discovered (see Figure F6). While Dirichlet regression and ANCOM-BC struggled with FDR control in all scenarios (see Figure F6), CompDA produced much higher FDR values for the abundant cell type (T4). For smaller sample sizes ($n = 100$) and small effects, *cosmoDA* was not able to consistently detect the differentially abundant features, resulting in lower power for these scenarios. With increasing sample size, the power of *cosmoDA* was on par with the other methods (Figures 4c, F7).

In the large-dimensional case ($p = 99$), the performance of all methods decreased in the small sample size scenario, while

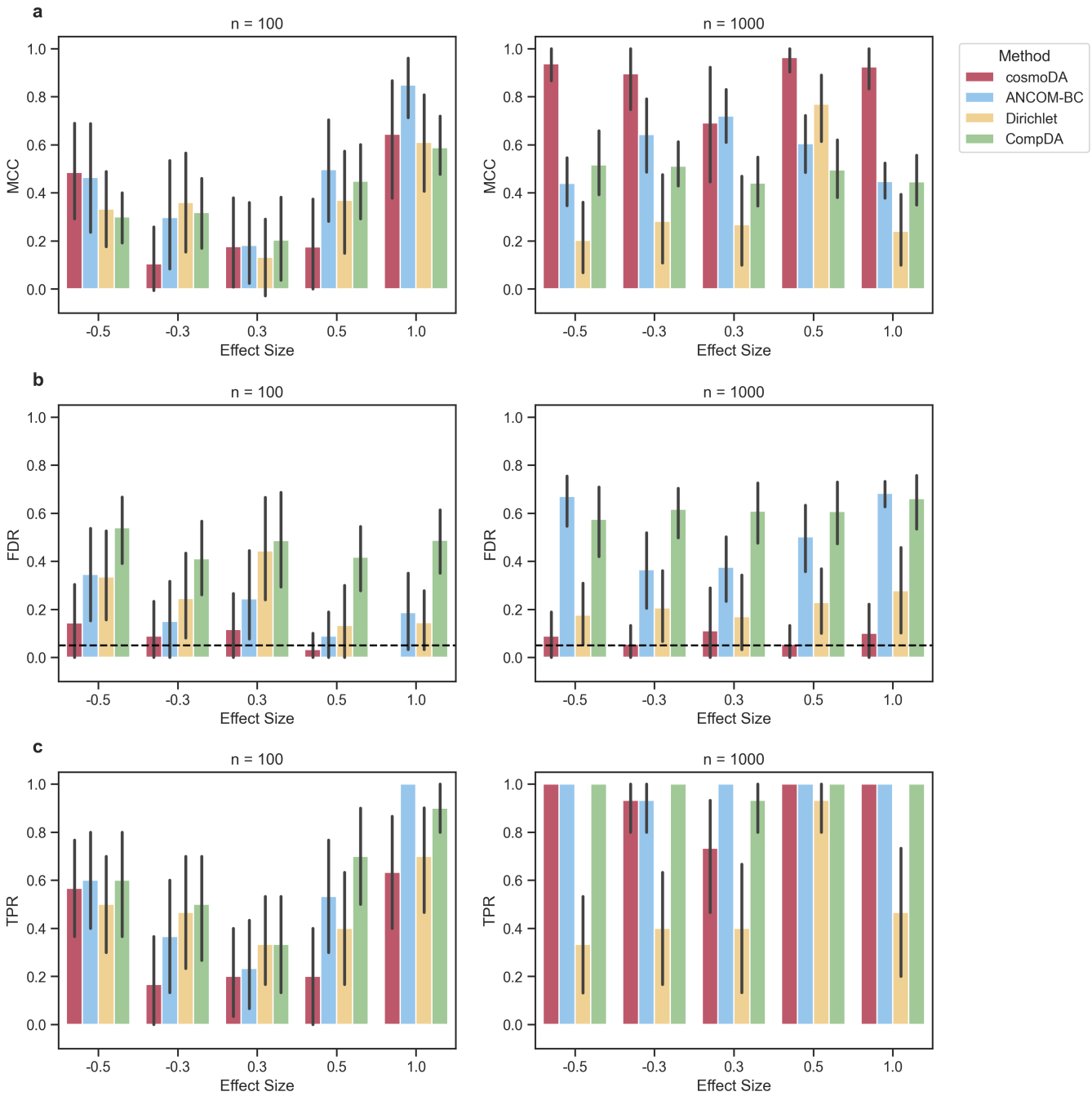


FIGURE 4 | Performance comparison for recovering differentially abundant features across different scenarios, $p = 11$. (a) Matthews' correlation coefficient, (b) False discovery rate. The dashed line shows the nominal FDR for all methods, and (c) True positive rate (power).

Matthews' correlation coefficient was similar to the case of $p = 11$ for larger sample sizes (Figures 5a, F11). Again, *cosmoDA* always showed the lowest FDR, albeit with slightly elevated levels for $n = 1000$, and mean FDR levels between 0.1 and 0.4 for $n = 100$ (Figure 5b). The FDR levels for *cosmoDA* did not show a trend across feature rarity and effect size, while the other methods were not able to produce average FDR levels below 0.5 for effects on rare pDC cells (Figure F12). In terms of power, only ANCOM-BC and Dirichlet regression were able to correctly detect some differentially abundant features for $n = 100$, while all methods showed good power for larger sample sizes (Figure 5c). Breaking these results down by cell type revealed a good power of ANCOM-BC and Dirichlet regression for abundant features, while effects on

rare features could not be reliably detected by any method (Figure F13). The unsuitability of *cosmoDA* and *CompDA* for the case of $p = 99$, $n = 100$ is not surprising, as both models need to estimate pairwise feature interactions in the high-dimensional regime.

4 | Applications to Single-Cell and Microbiome Data

To showcase the DA testing capabilities of *cosmoDA* on real data, we considered two compositional datasets: PBMC abundances derived from scRNA-seq data of SLE patients (as used in the semi-synthetic benchmarks [45]) and infant gut microbiome

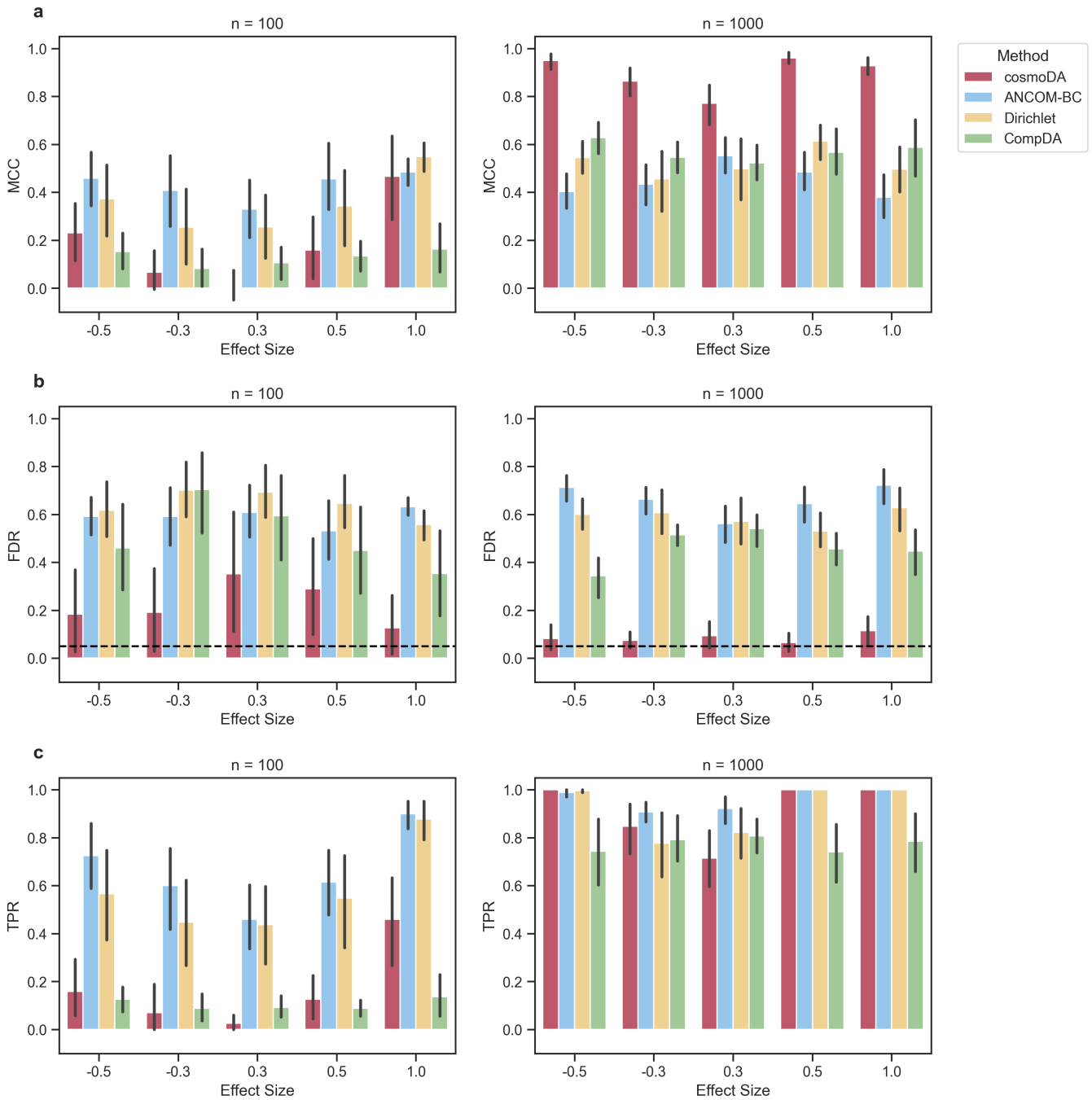


FIGURE 5 | Performance comparison for recovering differentially abundant features across different scenarios, $p = 99$. (a) Matthews' correlation coefficient, (b) False discovery rate. The dashed line shows the nominal FDR for all methods, and (c) True positive rate (power).

data from 16S rRNA sequencing [48]. Apart from comparing the empirical results with other state-of-the-art methods, we also evaluated the impact of power transformations ($a = b = \phi \neq 0$) on the downstream DA results.

4.1 | DA Analysis of Cell Type Compositions in Patients With Systemic Lupus Erythematosus

We used the scRNA-Seq-derived PBMC data from Reference [45] ($n = 352, p = 11$, see Section 3.2) to estimate differences in cell type composition between subjects with systemic lupus erythematosus (SLE) ($n = 208$) and healthy controls ($n = 144$).

To tune the parameters in `cosmoDA`, we considered the power values $\phi = (0.01, 0.02, 0.03, \dots, 0.99)$, as well as the log-log model ($\phi = 0$) for comparison. We set the range of λ_1 values between 1.5 and 10^{-7} to ensure full coverage of the range of supports of \mathbf{K} for every value of ϕ . As before, we used NK cells as the reference cell type for `cosmoDA` and selected the regularization strength via 5-fold cross-validation with the 1SE rule. We used ANCOM-BC, Dirichlet regression, and CompDA for comparison.

We first investigated the influence of our power value tuning schemes for DA analysis. The Procrustes correlation analysis showed that the ALR-transformed PBMC data (with zeros replaced by a pseudocount of 0.5) and the power-transformed

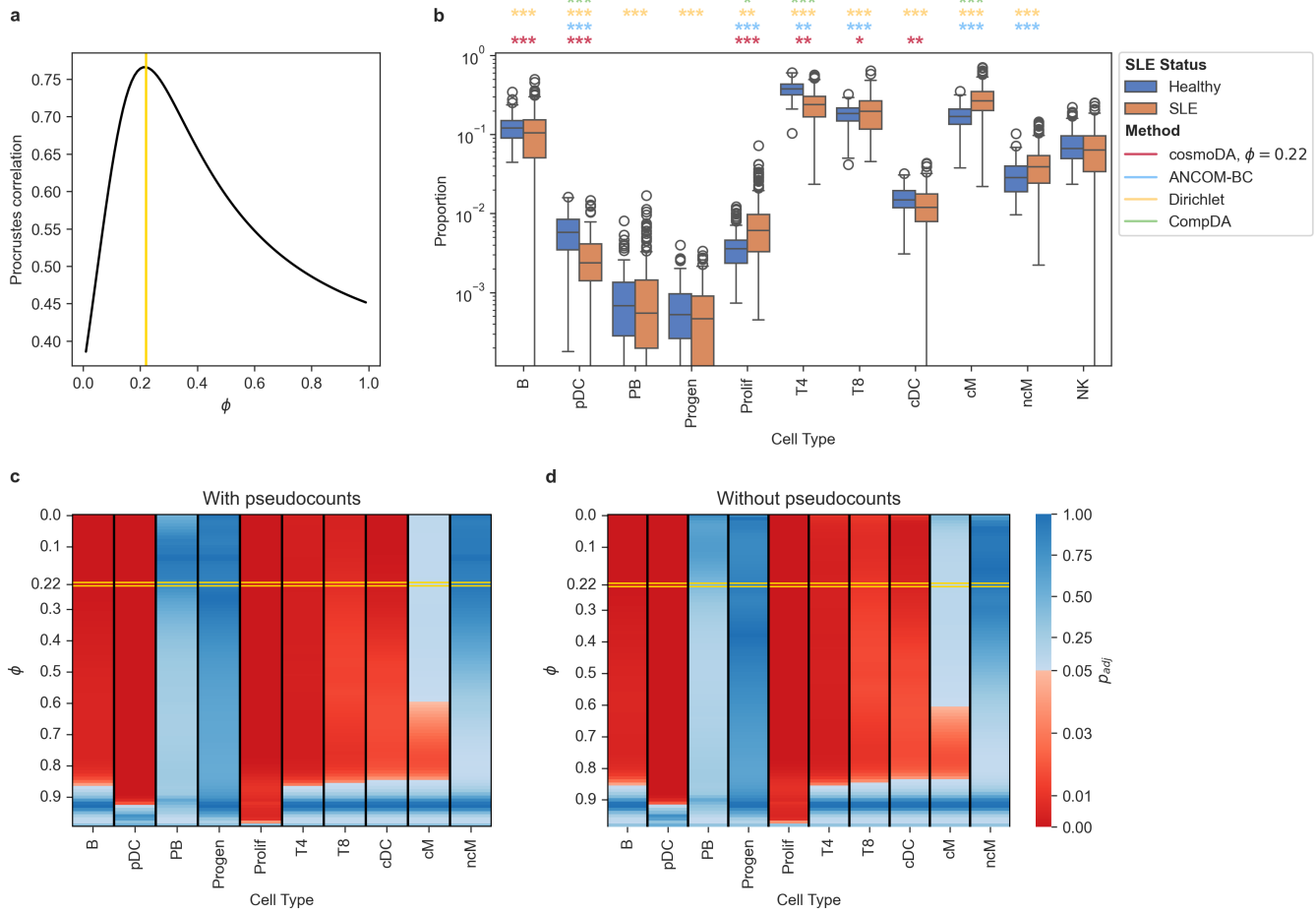


FIGURE 6 | Differential abundance testing with `cosmoDA` on the lupus dataset. (a) Procrustes correlation between power transformation and ALR transformation with zero replacement. The yellow line ($\phi^* = 0.22$) indicates the maximal Procrustes correlation. (b) Boxplot of relative abundance data without zero replacement. The colored stars indicate the significance level for each method (*: $p_{adj} < 0.05$; **: $p_{adj} < 0.01$; ***: $p_{adj} < 0.001$). Differential abundance results on NK cells (reference in `cosmoDA`) are omitted. (c) Adjusted p -values for testing differential abundance with `cosmoDA` on zero-replaced data with different power transformations. Red entries denote differential abundance at a level of $\alpha = 0.05$, blue entries denote no differential abundance. The yellow box highlights the adjusted p -values for ϕ^* determined in a. (d) Same as c, but using the raw data without zero replacement.

data had the highest alignment for a power of $\phi^* = 0.22$ (see Figure 6a). To investigate the impact of zero replacement and the power transform on differential abundance, we also compared the DA testing results of `cosmoDA` for all values of ϕ with and without zero entries (Figure 6c,d). Due to the low number of zeroes (4.5%), the impact of zero imputation was negligible for this dataset, making the adjusted p -values with and without zero imputation almost identical. Below a value of 0.8, the exponent of the power transformation only impacted the differential abundance of CD14+ classical monocytes (cM). For higher exponents, almost all cell types showed no differential abundance. Comparing the results of `cosmoDA` with ANCOM-BC, Dirichlet regression, and CompDA (all implemented as described in Section 3.2) showed that all methods selected different sets of differentially abundant cell types (Figure 6b). CompDA produced the most conservative results, only finding four DA cell types at an FDR level of 0.05. On the other hand, Dirichlet regression found all cell types to be differentially abundant. Interestingly, `cosmoDA` was the only method that did *not* select classical monocytes as differentially abundant (Figure 6b). The latter finding is in agreement

with a control experiment performed by [45] that found absolute monocyte abundances to be not differentially abundant in SLE patients.

4.2 | DA Analysis in Microbiome Data

To showcase the suitability of `cosmoDA` for microbial 16S rRNA sequencing, we used gut microbiome data from infants in Malawi and the United States [48]. We followed the pre-processing in the original ANCOM-BC study [26] and aggregated the data to the Phylum level. We selected all samples from subjects aged less than two years old in Malawi and the United States. Following [26], we next discarded all phyla where more than 90% of samples contained zero entries, resulting in $n = 97$ samples of $p = 13$ phyla. We selected Bacteroidetes as the reference phylum and applied `cosmoDA` with $\phi = (0.01, 0.02, 0.03, \dots, 0.99)$, and the log-log model ($\phi = 0$) to the relative abundances with and without zero replacement. The range of values for λ_1 was set to $[10^{-12}, 1.5]$, and we used 5-fold cross-validation with the 1SE rule to select λ_1 for each value of ϕ .

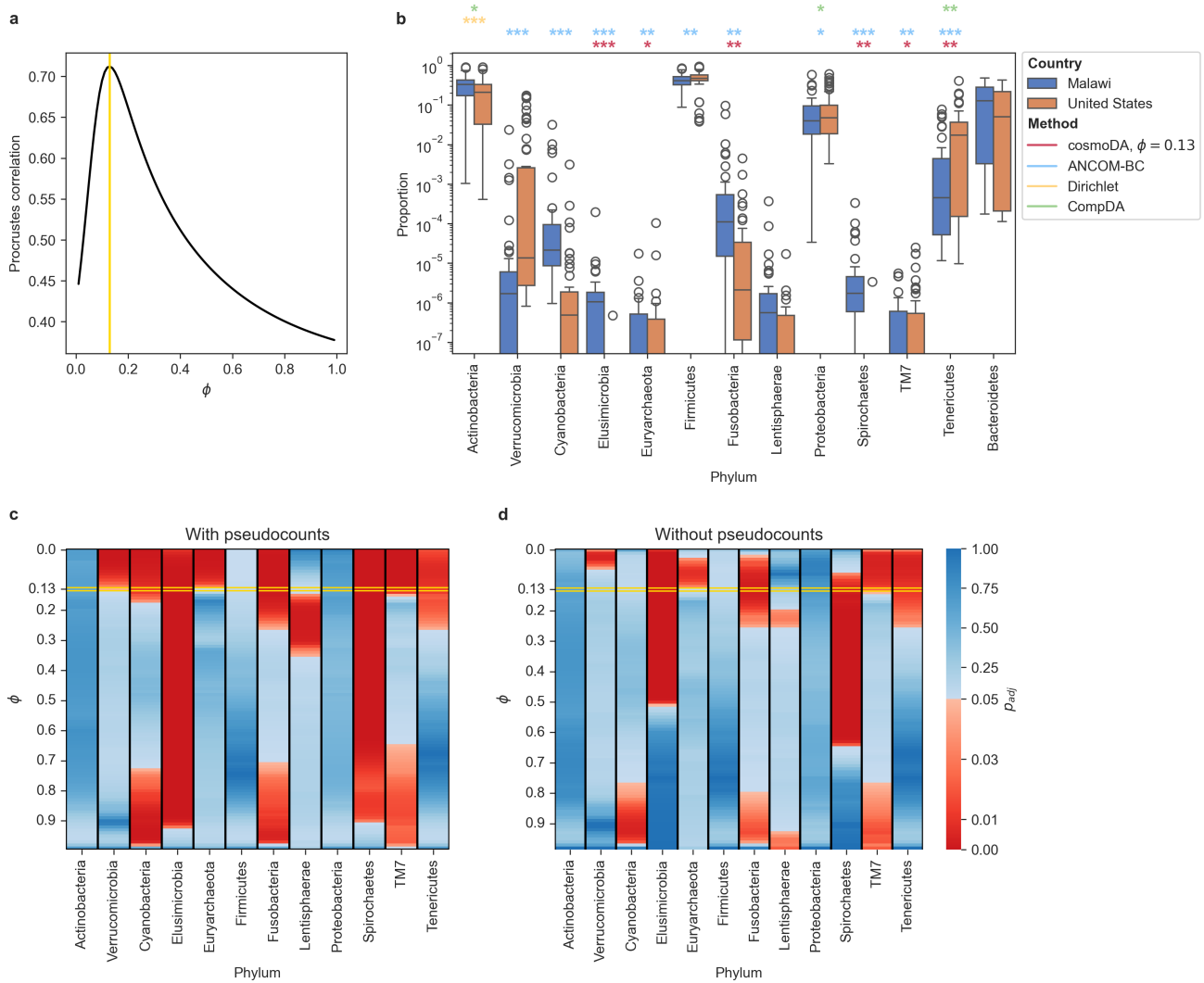


FIGURE 7 | Differential abundance testing (US vs. Malawi) with *cosmoDA* on infants (age < 2 years) in the human gut dataset. (a) Procrustes correlation between power transformation and ALR transformation with zero replacement. The yellow line ($\phi^* = 0.13$) indicates the maximal Procrustes correlation. (b) Boxplot of relative abundance data without zero replacement. The colored stars indicate the significance level for each method (*: $p_{adj} < 0.05$; **: $p_{adj} < 0.01$; ***: $p_{adj} < 0.001$). Differential abundance results on Bacteroidetes (reference in *cosmoDA*) are omitted. (c) Adjusted p -values for testing differential abundance with *cosmoDA* on zero-replaced data with different power transformations. Red entries denote differential abundance at a level of $\alpha = 0.05$, blue entries denote no differential abundance. The yellow box highlights the adjusted p -values for ϕ^* determined in a. (d) Same as c, but using the raw data without zero replacement.

For this dataset, our power selection scheme identified $\phi^* = 0.13$ to result in the best Procrustes alignment (see Figure 7a). The larger proportion of zero entries in this dataset (28.6%) caused more differences in downstream DA testing results, both on the original and zero-imputed data (Figure 7c,d). While the DA pattern of taxa with no zero entries (Firmicutes, Actinobacteria, Tenericutes, and Proteobacteria) was not impacted by zero imputation, the phyla with at least 20% zero entries (Cyanobacteria, Elusimicrobia, Euryarchaeota, Lentisphaerae, Spirochaetes, and TM7) were deemed differentially abundant at a smaller power value. Similar to the analysis of the scRNA-seq data, the four DA methods produced different sets of DA taxa at an FDR level of 0.05 (Figure 7b). Dirichlet regression and CompDA seemed to be only sensitive to taxa with high average abundance, while ANCOM-BC and *cosmoDA* were also able to detect differential abundance in rare phyla. The set of DA taxa discovered by *cosmoDA* at $\phi^* = 0.13$ on the data with

zero entries was smaller than the set discovered on the same dataset by ANCOM-BC. Nevertheless, *cosmoDA* found multiple phyla that are associated with rural lifestyles (Elusimicrobia, Euryarchaeota, Spirochaetes) to be increased in infants from Malawi [49, 50]. Notably, the ANCOM-BC algorithm involves the replacement of zeros by a small pseudocount [26]. Indeed, the set of DA phyla discovered by *cosmoDA* on the zero-replaced data with the exponent $\phi^* = 0.13$ (Figure 7c) almost perfectly matched the DA phyla found by ANCOM-BC (except Firmicutes and Proteobacteria). Overall, this confirms that the replacement of zero entries in microbial abundance data has a significant impact on differential abundance.

5 | Conclusion

Tissues and bacterial communities are complex biological environments, governed by interactions between individual cell types

or microbial taxa. The prevailing high-throughput sequencing (HTS) data sets probing these complex mixtures are often compositional in nature. Statistical generative modeling, as well as differential abundance testing schemes for such compositional datasets, can therefore suffer from inaccuracies if interactions between cells or microbes are not considered in the analysis. Extending the class of a-b power interaction models [24] by a linear effect on the location vector, our new method `cosmoDA` allows accurate modeling of HTS data with pairwise feature interactions in the presence of covariate information. The covariate formulation in `cosmoDA` also seamlessly integrates into the generalized score matching optimization framework [25, 35, 36], facilitating fast and accurate parameter inference. L_1 regularization on the interaction matrix further avoids model complexity explosion and allows to select parsimonious interaction patterns. Compared to the a-b power interaction model without covariates from Reference [24], the addition of a covariate did not reduce its ability to detect significant feature interactions in our synthetic data simulations. Both the covariate-less and covariate-extended a-b power interaction models outperformed other established procedures for identifying sparse interactions in compositional HTS data when the sample size was sufficiently large.

In the presence of a binary condition, testing for the significance of the covariate-related parameters in the location vector acts as a form of differential abundance testing. Here, the parallel estimation of feature interactions helps to avoid false positive detections, which are only indirectly related to the condition. In our realistic simulation experiments, `cosmoDA` was the only method to approximately control the false discovery rate in the presence of feature interactions, while no other tested method could distinguish between direct and indirect compositional changes. `cosmoDA` showed reduced power when the sample size was small, but was on par with methods like ANCOM-BC [26] for larger numbers of observations. One exception where `cosmoDA` was not able to adequately control the FDR was for misspecified models with more features than samples. We further demonstrated the ability of `cosmoDA` to find biologically meaningful differential abundances on two experimental datasets from human single-cell RNA sequencing and microbiome 16S rRNA sequencing.

The use of power transformations instead of the logarithm in a-b power interaction models allows to keep zero measurements in the data as is, avoiding distortions caused by imputation of these values. Through a small adjustment in the score matching optimizer, we were able to approximate the log-transformation for exponents approaching zero. Applying `cosmoDA` to real-world single-cell and microbiome datasets, we discovered that zero replacement and the exponent of the power transformation had a considerable impact on downstream DA results in data with excess zeros [8]. We further demonstrated that selecting an exponent for the power transformation that approximates the data geometry after an ALR transformation generally produces sensible differential abundance results.

While `cosmoDA` successfully tackles multiple challenges in generative modeling and differential abundance testing, it also has some limitations. Currently, `cosmoDA` can only accommodate a single binary or continuous covariate. Extending the linear model formulation would allow to model more complex

scenarios and adjusting for confounders in DA testing. For this, the score matching estimator would also have to be extended to multiple covariates. The implementation of such a flexible model could be simplified by using automatic differentiation for determining the elements of Γ and g [51]. In addition, we believe that approximation of the logarithm for small exponents can be solved more elegantly by changing the general definition of a-b power interaction models to utilize a true Box–Cox transformation rather than using our proposed adjustments in the score matching optimizer. Estimation of our model also relies on selecting a good reference, which is profiled out in the model formulation. Looping over multiple references and averaging the results, as described by [24], could avoid this dependency at the cost of computational efficiency.

While we empirically showed the feasibility of `cosmoDA`, we did not provide extensive formal guarantees for goodness of fit and convergence. Formally evaluating and extending our brief discussion of how the convergence criteria for a-b power interaction models provided by [24] are applicable to our model (Section 2.2.2) would give more justification to our approach. Similarly, a thorough goodness-of-fit comparison of our model and other approaches for modeling compositional high-throughput sequencing data (e.g., [3, 22, 32]) on a wide variety of real datasets could help characterize the strengths and weaknesses of different parametric and nonparametric approaches further. While such an evaluation is beyond the scope of this work, possible metrics for such a comparison could include the Fisher divergence, Kernelized Stein discrepancy [52], or simulation-based predictive comparisons.

Furthermore, `cosmoDA` requires any count data to be transformed to relative abundance data and treated as quasi-continuous. Due to this choice, inference of effects on very rare features can become imprecise due to the discrete lattice structure of count data [53]. In our approach, this effect can only be mitigated by conducting experiments with higher sequencing depth.

Overall, we believe that `cosmoDA`, with its abilities to include feature interactions and seamless handling of excess zeros, represents a valuable addition to the growing family of differential abundance testing methods. A Python implementation of `cosmoDA` and the power interaction model without covariates is available at <https://github.com/bio-datascience/cosmoDA>.

Author Contributions

J.O. developed the idea of `cosmoDA` with help from H.L. and C.L.M. J.O. implemented the method, performed and evaluated all simulations, and conducted the data applications. J.O. wrote the manuscript with assistance from C.L.M. All authors read and approved the final manuscript.

Acknowledgments

We thank Matthias Drton for his inputs to the score matching optimization in `cosmoDA`, and Fabian Schaipp for his recommendations regarding the optimization algorithm. We also thank Oleg Vlasovets for implementing the graphical lasso on the simulated benchmark data. This work was supported by the Helmholtz Association's Initiative and Networking Fund through Opt4Bio (Grant No. ZT-I-PF-5-138), awarded to C.L.M. H.L. was supported by NIH (Grant No. R01GM 123056).

Funding

This work was supported by the Helmholtz-Gemeinschaft (Grant No. ZT-I-PF-5-138) and National Institutes of Health (Grant No. R01GM123056).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All datasets used in this article are publicly available. The SLE scRNA-seq data were downloaded from the Human Cell Atlas platform (GSE174188). The gut microbiome data is stored at MG-RAST <https://www.mg-rast.org/index.html> under search string “mgp401”, code for data preparation was adapted from <https://github.com/FrederickHuangLin/ANCOM-BC>. Code for reproducing the analyses in this article is available under <https://github.com/bio-datascience/cosmoDA>, intermediate data can be found at <https://zenodo.org/records/13911623>.

References

1. T. P. Quinn, I. Erb, M. F. Richardson, and T. M. Crowley, “Understanding Sequencing Data as Compositions: An Outlook and Review,” *Bioinformatics* 34, no. 16 (2018): 2870–2878, <https://doi.org/10.1093/bioinformatics/bty175>.
2. M. C. B. Tsilimigras and A. A. Fodor, “Compositional Data Analysis of the Microbiome: Fundamentals, Tools, and Challenges,” *Annals of Epidemiology* 26, no. 5 (2016): 330–335, <https://doi.org/10.1016/j.annepidem.2016.03.002>.
3. M. Büttner, J. Ostner, C. L. Müller, F. J. Theis, and B. Schubert, “scCODA Is a Bayesian Model for Compositional Single-Cell Data Analysis,” *Nature Communications* 12, no. 1 (2021): 6876, <https://doi.org/10.1038/s41467-021-27150-6>.
4. L. Heumos, A. C. Schaar, C. Lance, et al., “Best Practices for Single-Cell Analysis Across Modalities,” *Nature Reviews Genetics* 24 (2023): 1–23, <https://doi.org/10.1038/s41576-023-00586-w>.
5. A. Regev, S. A. Teichmann, E. S. Lander, et al., “The Human Cell Atlas,” *eLife* 6 (2017): 6, <https://doi.org/10.7554/eLife.27041>.
6. R. McNulty, D. Sritharan, S. H. Pahnig, et al., “Probe-Based Bacterial Single-Cell RNA Sequencing Predicts Toxin Regulation,” *Nature Microbiology* 8, no. 5 (2023): 934–945, <https://doi.org/10.1038/s41564-023-01348-4>.
7. P. J. Turnbaugh, R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon, “The Human Microbiome Project,” *Nature* 449, no. 7164 (2007): 804–810, <https://doi.org/10.1038/nature06244>.
8. G. B. Gloor, J. M. Macklaim, V. Pawlowsky-Glahn, and J. J. Egozcue, “Microbiome Datasets Are Compositional: And This Is Not Optional,” *Frontiers in Microbiology* 8 (2017): 2224, <https://doi.org/10.3389/fmicb.2017.02224>.
9. J. Aitchison, “The Statistical Analysis of Compositional Data,” *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 44, no. 2 (1982): 139–160.
10. M. I. Love, W. Huber, and S. Anders, “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data With DESeq2,” *Genome Biology* 15, no. 12 (2014): 550, <https://doi.org/10.1186/s13059-014-0550-8>.
11. A. K. Mishra and C. L. Müller, “Negative Binomial Factor Regression With Application to Microbiome Data Analysis,” *Statistics in Medicine* 41, no. 15 (2022): 2786–2803, <https://doi.org/10.1002/sim.9384>.
12. R. H. Hijazi and R. W. Jernigan, “Modelling Compositional Data Using Dirichlet Regression Models,” *Journal of Applied Probability & Statistics* 4, no. 1 (2009): 77–91.
13. W. D. Wadsworth, R. Argiento, M. Guindani, J. Galloway-Pena, S. A. Shelburne, and M. Vannucci, “An Integrative Bayesian Dirichlet-Multinomial Regression Model for the Analysis of Taxonomic Abundances in Microbiome Data,” *BMC Bioinformatics* 18, no. 1 (2017): 94, <https://doi.org/10.1186/s12859-017-1516-0>.
14. J. Ostner, S. Carcy, and C. L. Müller, “tascCODA: Bayesian Tree-Aggregated Analysis of Compositional Amplicon and Single-Cell Data,” *Frontiers in Genetics* 12 (2021): 766405, <https://doi.org/10.3389/fgene.2021.766405>.
15. J. Aitchison and S. M. Shen, “Logistic-Normal Distributions: Some Properties and Uses,” *Biometrika* 67, no. 2 (1980): 261–272, <https://doi.org/10.2307/2335470>.
16. F. Xia, J. Chen, W. K. Fung, and H. Li, “A Logistic Normal Multinomial Regression Model for Microbiome Compositional Data Analysis,” *Biometrics* 69, no. 4 (2013): 1053–1063, <https://doi.org/10.1111/biom.12079>.
17. Y. Zeng, D. Pang, H. Zhao, and T. Wang, “A Zero-Inflated Logistic Normal Multinomial Model for Extracting Microbial Compositions,” *Journal of the American Statistical Association* 544 (2022): 2356–2369, <https://doi.org/10.1080/01621459.2022.2044827>.
18. J. D. Silverman, K. Roche, Z. C. Holmes, L. A. David, and S. Mukherjee, “Bayesian Multinomial Logistic Normal Models Through Marginally Latent Matrix-T Processes,” *Journal of Machine Learning Research* 23, no. 7 (2022): 1–42.
19. S. Lubbe, P. Filzmoser, and M. Templ, “Comparison of Zero Replacement Strategies for Compositional Data With Large Numbers of Zeros,” *Chemometrics and Intelligent Laboratory Systems* 210 (2021): 104248, <https://doi.org/10.1016/j.chemolab.2021.104248>.
20. M. Greenacre, E. Grunsky, J. Bacon-Shone, I. Erb, and T. Quinn, “Aitchison’s Compositional Data Analysis 40 Years on: A Reappraisal,” *SSO Schweizerische Monatsschrift Fur Zahnheilkunde* 38, no. 3 (2023): 386–410, <https://doi.org/10.1214/22-STSS880>.
21. J. Aitchison, “A General Class of Distributions on the Simplex,” *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 47, no. 1 (1985): 136–146, <https://doi.org/10.1111/j.2517-6161.1985.tb01341.x>.
22. J. L. Scaely and A. T. A. Wood, “Score Matching for Compositional Distributions,” *Journal of the American Statistical Association* 118, no. 543 (2022): 1811–1823, <https://doi.org/10.1080/01621459.2021.2016422>.
23. J. L. Scaely, K. L. Hingee, J. T. Kent, and A. T. A. Wood, “Robust Score Matching for Compositional Data,” *Statistics and Computing* 34, no. 2 (2024): 93, <https://doi.org/10.1007/s11222-024-10412-w>.
24. S. Yu, M. Drton, and A. Shojaie, “Interaction Models and Generalized Score Matching for Compositional Data,” in *Proceedings of the Second Learning on Graphs Conference. 231 of Proceedings of Machine Learning Research*, ed. S. Villar and B. Chamberlain (PMLR, 2024), 20:1–20:25.
25. A. Hyvärinen, “Estimation of Non-Normalized Statistical Models by Score Matching,” *Journal of Machine Learning Research* 6 (2005): 695–709.
26. H. Lin and S. D. Peddada, “Analysis of Compositions of Microbiomes With Bias Correction,” *Nature Communications* 11, no. 1 (2020): 3514, <https://doi.org/10.1038/s41467-020-17041-7>.
27. H. Zhou, K. He, J. Chen, and X. Zhang, “LinDA: Linear Models for Differential Abundance Analysis of Microbiome Compositional Data,” *Genome Biology* 23, no. 1 (2022): 95, <https://doi.org/10.1186/s13059-022-02655-5>.
28. J. T. Nearing, G. M. Douglas, M. G. Hayes, et al., “Microbiome Differential Abundance Methods Produce Different Results Across 38 Datasets,” *Nature Communications* 13, no. 1 (2022): 342.
29. S. Ma, C. Huttenhower, and L. Janson, “Compositional Differential Abundance Testing: Defining and Finding a New Type of Health-Microbiome Associations. bioRxiv 2024:2024.06.04.596112,” <https://doi.org/10.1101/2024.06.04.596112>.
30. D. Billheimer, P. Guttorp, and W. Fagan, “Statistical Interpretation of Species Composition,” *Journal of the American Statistical*

- Association 96, no. 456 (2001): 1205–1214, <https://doi.org/10.1198/016214501753381850>.
31. I. Erb, “Partial Correlations in Compositional Data Analysis,” *Applied Computing and Geosciences* 6 (2020): 100026.
32. C. Weistuch, J. Zhu, J. O. Deasy, and A. R. Tannenbaum, “The Maximum Entropy Principle for Compositional Data,” *BMC Bioinformatics* 23, no. 1 (2022): 449, <https://doi.org/10.1186/s12859-022-05007-z>.
33. A. Hyvärinen, “Some Extensions of Score Matching,” *Computational Statistics and Data Analysis* 51, no. 5 (2007): 2499–2512, <https://doi.org/10.1016/j.csda.2006.09.003>.
34. S. Yu, M. Drton, and A. Shojaie, “Generalized Score Matching for Non-Negative Data,” *Journal of Machine Learning Research* 20 (2019) 1–70, <https://jmlr.org/papers/v20/18-278.html>.
35. L. Lin, M. Drton, and A. Shojaie, “Estimation of High-Dimensional Graphical Models Using Regularized Score Matching,” *Electronic Journal of Statistics* 10, no. 1 (2016): 806–854, <https://doi.org/10.1214/16-EJS1126>.
36. S. Yu, M. Drton, and A. Shojaie, “Generalized Score Matching for General Domains,” *Information and Inference* 11, no. 2 (2022): 739–780, <https://doi.org/10.1093/imaia/iaaa041>.
37. Z. D. Kurtz, C. L. Müller, E. R. Miraldi, D. R. Littman, M. J. Blaser, and R. A. Bonneau, “Sparse and Compositionally Robust Inference of Microbial Ecological Networks,” *PLoS Computational Biology* 11, no. 5 (2015): e1004226, <https://doi.org/10.1371/journal.pcbi.1004226>.
38. Y. Benjamini and Y. Hochberg, “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing,” *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 57, no. 1 (1995): 289–300.
39. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, 2nd ed. (Springer, 2009).
40. R. Foygel and M. Drton, “Extended Bayesian Information Criteria for Gaussian Graphical Models,” *Advances in Neural Information Processing Systems* (2010): 23.
41. D. E. Te Beest, E. H. Nijhuis, T. W. R. Möhlmann, and C. J. F. Ter Braak, “Log-Ratio Analysis of Microbiome Data With Many Zeroes Is Library Size Dependent,” *Molecular Ecology Resources* 21, no. 6 (2021): 1866–1874, <https://doi.org/10.1111/1755-0998.13391>.
42. G. E. P. Box and D. R. Cox, “An Analysis of Transformations,” *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 26, no. 2 (1964): 211–243, <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>.
43. M. Greenacre, “The chiPower Transformation: A Valid Alternative to Logratio Transformations in Compositional Data Analysis,” *Advances in Data Analysis and Classification* 18 (2024): 769–796, <https://doi.org/10.1007/s11634-024-00600-x>.
44. M. Tsagris, S. Preston, and A. T. A. Wood, “Improved Classification for Compositional Data Using the α -Transformation,” *Journal of Classification* 33, no. 2 (2016): 243–261, <https://doi.org/10.1007/s00357-016-9207-5>.
45. R. K. Perez, M. G. Gordon, M. Subramaniam, et al., “Single-Cell RNA-Seq Reveals Cell Type-Specific Molecular and Genetic Associations to Lupus,” *Science* 376, no. 6589 (2022): eabf1970, <https://doi.org/10.1126/science.abf1970>.
46. F. Schaipp, O. Vlasovets, and C. Müller, “GGLasso - a Python Package for General Graphical Lasso Computation,” *Journal of Open Source Software* 6, no. 68 (2021): 3865, <https://doi.org/10.21105/joss.03865>.
47. M. J. Maier, *DirichletReg: Dirichlet Regression for Compositional Data in R. Research Report Series*, vol. 125 (Vienna University of Economics and Business, 2014), 125.
48. T. Yatsunencko, F. E. Rey, M. J. Manary, et al., “Human Gut Microbiome Viewed Across Age and Geography,” *Nature* 486, no. 7402 (2012): 222–227, <https://doi.org/10.1038/nature11053>.
49. D. P. R. Herlemann, O. Geissinger, and A. Brune, “The Termite Group I Phylum Is Highly Diverse and Widespread in the Environment,” *Applied and Environmental Microbiology* 73, no. 20 (2007): 6682–6685, <https://doi.org/10.1128/AEM.00712-07>.
50. A. J. Obregon-Tito, R. Y. Tito, J. Metcalf, et al., “Subsistence Strategies in Traditional Societies Distinguish Gut Microbiomes,” *Nature Communications* 6, no. 1 (2015): 6505, <https://doi.org/10.1038/ncomms7505>.
51. L. H. Kassel, J. Scealy, and B. M. Bell, “Score Matching Estimation by Automatic Differentiation,” (2024).
52. Q. Liu, J. Lee, and M. Jordan, “A Kernelized Stein Discrepancy for Goodness-Of-Fit Tests,” in *Proceedings of the 33rd International Conference on Machine Learning. 48 of Proceedings of Machine Learning Research*, ed. M. F. Balcan and K. Q. Weinberger (PMLR, 2016), 276–284.
53. D. R. Lovell, X. Y. Chua, and A. McGrath, “Counts: An Outstanding Challenge for Log-Ratio Analysis of Compositional Data in the Molecular Biosciences,” *NAR Genomics and Bioinformatics* 2, no. 2 (2020): lqaa040, <https://doi.org/10.1093/nargab/lqaa040>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** Online Appendix.