

Article

Toward Precision Radiotherapy: A Nonlinear Optimization Framework and an Accelerated Machine Learning Algorithm for the Deconvolution of Tumor-Infiltrating Immune Cells

Lois Chinwendu Okereke ^{1,*} , Abdulmalik Usman Bello ^{1,2} and Emmanuel Akwari Onwukwe ^{3,4}

¹ Department of Pure and Applied Mathematics, Mathematics Institute (Emerging Regional Centre of Excellence (ERCE) of the European Mathematical Society (EMS)), African University of Science and Technology, Abuja 900107, Nigeria

² Department of Mathematics, Federal University Dutsin-Ma, Dutsin-Ma 821101, Nigeria

³ Department of Theoretical and Applied Physics, African University of Science and Technology, Abuja 900107, Nigeria

⁴ Inspired Innovative Sustainable (IIS) Projects & Solutions Limited, Abuja 900107, Nigeria

* Correspondence: lokereke@aust.edu.ng

Abstract: Tumor-infiltrating immune cells (TIICs) form a critical part of the ecosystem surrounding a cancerous tumor. Recent advances in radiobiology have shown that, in addition to damaging cancerous cells, radiotherapy drives the upregulation of immunosuppressive and immunostimulatory TIICs, which in turn impacts treatment response. Quantifying TIICs in tumor samples could form an important predictive biomarker guiding patient stratification and the design of radiotherapy regimens and combined immune-radiation treatments. As a result of several limitations associated with experimental methods for quantifying TIICs and the availability of extensive gene sequencing data, deconvolution-based computational methods have appeared as a suitable alternative for quantifying TIICs. Accordingly, we introduce and discuss a nonlinear regression approach (remarkably different from the traditional linear modeling approach of current deconvolution-based methods) and a machine learning algorithm for approximating the solution of the resulting constrained optimization problem. This way, the deconvolution problem is treated naturally, given that the gene expression levels of pure and heterogenous samples do not have a strictly linear relationship. When applied across transcriptomics datasets, our approach, which also allows the coupling of different loss functions, yields results that closely match ground-truth values from experimental methods and exhibits superior performance over popular deconvolution-based methods.

Keywords: predictive biomarkers; bulk RNA-seq; nonlinear regression; inverse problem; digital cytometry; bioinformatics; immune contexture; nonlinear functional analysis; constrained optimization; error analysis



Citation: Okereke, L.C.; Bello, A.U.; Onwukwe, E.A. Toward Precision Radiotherapy: A Nonlinear Optimization Framework and an Accelerated Machine Learning Algorithm for the Deconvolution of Tumor-Infiltrating Immune Cells. *Cells* **2022**, *11*, 3604. <https://doi.org/10.3390/cells11223604>

Academic Editor:
Alexander E. Kalyuzhny

Received: 11 October 2022

Accepted: 9 November 2022

Published: 14 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent times, the continuous rise in the global cancer burden has emphasized the need for increased efforts in cancer treatment strategies. In 2020 alone, there were 19.3 million new cases and almost 10 million cancer deaths worldwide, with the number of new cases projected to climb to 28.4 million by 2040 [1]. This projection is not far-fetched, given the 36.9% increase in the number of new cases from 2012 to 2020 [2]. With up to half of all these cancer cases receiving radiotherapy during their treatment, radiotherapy is still a vital cancer treatment strategy [3–5].

In simple terms, radiotherapy involves using ionizing radiation, usually high-energy X-rays, to kill cancer cells or, at least, limit their proliferation by damaging their genetic code of life, known as deoxyribonucleic acid (DNA). In doing this, the goal is to achieve tumor control without introducing severe damage to surrounding normal tissues, enhancing

treatment outcomes and minimizing adverse effects. Precision radiotherapy aims to reach this goal by stratifying and precisely treating “each individual cancer patient, using state-of-the-art new radiotherapy technology and biomarkers” [6].

Biomarkers are objectively evaluated and measured characteristics indicative of normal (or abnormal) biological processes, pathogenesis, or therapeutic response [7]. Their roles could be prognostic, diagnostic, treatment response monitoring, or predictive [8]. In their predictive role, biomarkers indicate the likelihood of a therapeutic benefit from a specific treatment for a given patient. Thus, in the case of precision radiotherapy, the complementary role of biomarkers for predictive purposes evolved from recent findings, showing that the effects of radiotherapy on the tumor microenvironment (TME) may be a significant determinant of the efficacy of a radiotherapy regimen [8–10]. For instance, in addition to damaging the malignant part of the TME, radiotherapy has been found to trigger immunomodulatory effects and alterations to critical components of the TME, such as tumor-infiltrating immune cells (TIICs) [9,11–13]. The latter is of particular interest because, as radiotherapy drives the upregulation of immunostimulatory TIICs such as cytotoxic CD8⁺ T cells, and immunosuppressive TIICs such as regulatory T cells (Treg), its impact is felt on differing cell subsets [13]. Consequently, quantifying TIICs in pretreatment and treatment of tumor samples is crucial in identifying predictive biomarkers guiding patient stratification and designing suitable radiotherapy regimens, including combined immune-radiation treatments [14].

Traditionally, experimental methods such as immunohistochemistry (IHC), cytometry, and recently, single-cell RNA sequencing (scRNA-seq) have been the gold standard for quantifying TIICs in samples. Although these methods precisely quantify TIICs in samples, there are limitations in terms of the technicality and range of applicability associated with each method. On the one hand, scRNA-seq is not just expensive and laborious for routine use but also highly prone to bias due to variations in the dissociation efficiencies of single cells [15]. On the other hand, IHC and cytometry rely on a small number of phenotypic markers, exhibit low to medium throughput, have little or no public datasets available, and are difficult to apply in large tumor series [16]. This situation necessitates the search for suitable alternatives to quantify TIICs in tumor samples.

Recently, the sharp reduction in the cost of next-generation sequencing (NGS) technologies has encouraged its routine application in clinical settings, resulting in the availability of large amounts of transcriptomics datasets from patients’ tumor samples, such as The Cancer Genome Atlas (TCGA) [17]. Although these datasets represent the bulk tumor sample, they provide a suitable alternative for quantifying the sample TIICs using computational techniques [18]. Computational techniques serving this purpose are broadly categorized into two. The first broad category is marker gene-based methods [14,19]. Methods under this category utilize a list of genes characteristic of a cell type (called marker genes), to quantify every cell type independently from the expression levels of the marker genes in the heterogeneous tumor sample. As a result, marker gene-based methods can only generate “a semi-quantitative score describing the enrichment of a cell type in a sample” [14], thus effectively making the comparison between cell types impractical.

The second broad category, deconvolution-based methods, considers the gene expression profile of the heterogeneous tumor sample as a *convolution* of the gene expression levels of the different cell components [20]; as a result, they can quantitatively estimate the fractions of cell types of interest (in this case TIICs). This consideration allows the problem to be formulated mathematically as a function of the gene expression profiles of the cell-type admixture. Thus, given the bulk gene expression of a tumor sample and a known cell-type specific expression profile, solving an inverse problem can estimate the cell-type fractions in the heterogeneous tumor sample.

To date, most deconvolution-based methods, including those specific to quantifying TIICs, such as CIBERSORT [21], CIBERSORTx [22], EPIC [23], ECIS [24], quanTIseq [25], and TIMER [26], assume that function to be linear. However, they yield different results for different cell types and use cases despite utilizing the gradient algorithm or its variants to

approximate solutions to the inverse problem. Interestingly, this is because each method is conceptually different according to the choice of loss function (or objective functions) and the setting of the optimization problem (constrained or unconstrained). It is against this backdrop that packages such as Immunedecon [27], TIMER2.0 [28], and TumorDecon [29] have sought to provide a unified platform that allows each of the different methods to be applied on the same dataset to compare or complement results. Accordingly, the strengths of each method can be harnessed to gain more robust and comprehensive estimates. Nevertheless, this approach is still susceptible to the potential issues associated with traditional linear modeling [30], given that “the relationship between the expression levels of pure and heterogeneous samples is not strictly linear” [14]. Moreover, dealing with large transcriptomics datasets calls for computationally efficient methods with fast rates of convergence and runtime [30].

Therefore, the main aim of this paper is to introduce and discuss a mathematical formulation that permits the TIIC deconvolution inverse problem to be handled in its natural state alongside an accelerated machine learning algorithm for approximating its solution. Through rigorous mathematical analysis, we show that the algorithm converges to an optimal solution of the inverse problem for various loss functions. More specifically, a globally optimal solution is guaranteed for convex loss functions. Furthermore, we use numerical experiments to show that the algorithm exhibits faster convergence rates and runtime than other traditionally used algorithms. When applied across transcriptomics datasets, our results closely match values from experimental methods and show superior performance over popular TIIC deconvolution-based methods. We end with a note on the detailed science behind these observations and an explanation of how this framework can be applied across similar inverse problems in biology, medical physics, and oncology.

2. Materials and Methods

2.1. Formulation and Discussion of the Deconvolution Problem

Let N denote the number of different cell types forming a mixture sample, and M be the number of genes whose expressions are measured in the sample. Let $B = (b_1, b_2, \dots, b_M) \in \mathbb{R}^M$ be the measurements of gene expression in the sample. Let $S \in \mathbb{R}^{M \times N}$ be the corresponding reference expressions matrix of the M genes from the N constituent cell types, and $P = (p_1, p_2, \dots, p_N) \in \mathbb{R}^N$ be the unknown proportions of mix of the different cell types. An operator r can model the relationship of B , S , and P as

$$B = r(S, P). \quad (1)$$

The deconvolution problem is concerned with the inverse problem of estimating P , given B and S . This inverse problem can be formulated mathematically as the following equivalent constrained optimization problem:

$$\min_{P \in C} \mathcal{L}(B, r(S, P)), \quad (2)$$

where $\mathcal{L}$ is a loss function measuring model fitness, and C is the set of constraints on P arising naturally from its definition as proportions, i.e.,

$$C = \left\{ P \in \mathbb{R}^N : p_i \geq 0 \forall i = 1, \dots, N, \sum_{i=1}^N p_i = 1 \right\}.$$

Many of the existing deconvolution methods (see, for example, [21–23,31–35] and references therein) consider B , S , and P to be linearly related in the form

$$B = SP + e, \quad (3)$$

where e is a random error. Several issues with the linear framework have been identified [30]. More recently, the authors of [36] showed problems associated with different

scales of gene expression within the linear framework and then proposed the following hybrid model:

$$\log(b_i) = d + \log\left(\sum_{j=1}^N S_{ij}p_j\right) + e_i, \quad (4)$$

where d accounts for systemic technical variation. Equivalently,

$$B \approx \exp(d)SP. \quad (5)$$

We note that $\exp(d)$ in Equation (5) is a constant factor adjustment across all genes. However, more than this constant factor adjustment may be required, because such systemic technical variations affect genes differently [37]. Thus, one may consider a more general model of the form

$$B \approx DSP, \quad (6)$$

where D is a diagonal matrix with diagonal entries as gene-specific factor adjustment generated from some known distribution. Accordingly, Equations (3) and (5) become special cases of Equation (6). Nevertheless, such a linear transformation may not efficiently describe nonlinear patterns.

Consequently, we formulate a generic nonlinear framework for the deconvolution problem by considering the operator r in Equation (1) as a nonlinear transformation involving S and P . This is because generalized nonlinear regression methods have been shown to yield better prediction accuracy for complex nonlinear patterns, where traditional linear regression models may fail [38]. Specifically, our nonlinear operator r is given as

$$r_i = \left(\delta + \sum_{j=1}^N S_{ij}p_j\right)^\theta, \quad \delta, \theta > 0. \quad (7)$$

The nonlinearity of Equation (7) depends on the value of θ . Note that Equation (7) reduces to the linear framework for $\theta = 1$. Therefore, for an arbitrary sequencing dataset, it is worthwhile to determine what values of θ (different from one) closely describe the not strictly linear relationship of the gene expression profiles of pure and heterogeneous samples.

Rigorous data assimilation techniques are useful in making such determinations from arbitrary datasets. However, we refrain from such nontrivial rigorous analytical examinations as they are beyond the aims of this work. As a result, we choose our θ values using an empirical approach for the purpose of demonstrating the proof of principle which is the subject of this work. The empirical approach is based on our hypothesis that, for some $\theta \in (0, 1) \cup (1, 2)$, we may be able to get suitable values satisfying the description of the relationship between the gene expression profiles of pure and heterogeneous samples. More specifically, we hypothesize that such a value might be slightly less than one or slightly greater than one on the order of a few decimal places. This hypothesis is guided by our preferred interpretation of the expression “not strictly linear”.

2.2. Review of Some Commonly Used Loss Functions

Several research surveys of cell-type deconvolution methods (see, for example, [14,19,39]) have identified the quadratic/squared error loss and the ε -insensitive loss, as the most commonly used loss functions for reference-based cell-type deconvolution within the linear framework. On the one hand, the squared error loss (SEL) is formulated on the basis of squared deviations as follows:

$$\mathcal{L}(B, r(S, P)) = \sum_{i=1}^M (b_i - S_i P)^2 = B - SP_2^2, \quad (8)$$

where S_i denotes the i -th row of S . In fact, it is the most common choice of loss function due to its simplicity. Notably, the SEL is highly susceptible to outliers. This feature can

be especially beneficial when the outliers originate naturally from variations within the process and, as such, contain useful systemic information. Conversely, it can be a drawback when the outliers arise from noise (errors).

On the other hand, the ε -insensitive loss (also referred to as the support vector method) [40] and other robust techniques, such as the Huber and Laplacian losses, are employed to reduce the drawback of the SEL. A unified version of these robust techniques was introduced in [41] as a soft insensitive loss function (SILF), expressed mathematically as

$$\mathcal{L}(B, r(S, P)) = \sum_{i=1}^M \uparrow_i(b_i - S_i P), \quad (9)$$

where

$$\uparrow_i(a) = \begin{cases} -a - \varepsilon, & \text{if } a \in (-\infty, -(1 + \rho)\varepsilon) \\ \frac{(a + (1 - \rho)\varepsilon)^2}{4\rho\varepsilon}, & \text{if } a \in [-(1 + \rho)\varepsilon, -(1 - \rho)\varepsilon] \\ 0, & \text{if } a \in (-(1 - \rho)\varepsilon, (1 - \rho)\varepsilon) \\ \frac{(a - (1 - \rho)\varepsilon)^2}{4\rho\varepsilon}, & \text{if } a \in [(1 - \rho)\varepsilon, (1 + \rho)\varepsilon] \\ a - \varepsilon, & \text{if } a \in ((1 + \rho)\varepsilon, +\infty) \end{cases},$$

with $0 < \rho \leq 1$ and $\varepsilon > 0$. In [41], the authors remarked that this function (Equation (9)) is smooth and inherits most of the desirable characteristics of several robust techniques, including insensitivity to outliers. They further demonstrated in great detail the computational efficiency and competitiveness of SILF compared to other well-respected techniques. Remarkably, these loss functions have been a dominating paradigm in the deconvolution problem, mainly due to their convexity.

Nevertheless, it has been shown in recent times that nonconvex loss functions improve the generic applicability and robustness of learning, especially in situations where the data and noise distributions are unknown [42]. One such loss function is given in Equation (10) as a Cauchy kernel risk-sensitive loss (CKRSL), derived using a Gaussian kernel-adapted operator and following methods similar to those in [43,44].

$$\mathcal{L}(B, r(S, P)) = \frac{1}{M} \sum_{i=1}^M \beta \log \left[1 + 2 \left(\frac{1 - \exp \left(- \left(\frac{\sum_{j=1}^N S_{ij} p_j - b_i}{2\sigma} \right)^2 \right)}{\beta} \right) \right], \quad \sigma > 0. \quad (10)$$

We remark that the abstract formulation presented in Section 2.1 has the advantage of accommodating several loss functions including those reviewed in this subsection.

2.3. Specification of the Loss Function for this Study

For the present study, we use a weighted squared error loss (WSEL) on the relationship operator r defined by Equation (7), expressed as

$$\mathcal{L}(B, r(S, P)) = \frac{1}{M} \sum_{i=1}^M \left(\left(\delta + \sum_{j=1}^N S_{ij} p_j \right)^\theta - b_i \right)^2. \quad (11)$$

2.4. Accelerated Machine Learning Algorithm (AMLA)

An optimal solution to the constrained optimization problem of Equation (2) can be approximated by the following accelerated machine learning algorithm (AMLA):

$$\begin{cases} v_0, v_1 \in C \\ w_j = v_j + \alpha_j (v_{j-1} - v_j), \\ v_{j+1} = \wp_C(w_j - \lambda A v_j) \end{cases} \quad (12)$$

for some predefined values of the adaptive momentum parameter α_j and step size (or learning rate) λ . A denotes the gradient of the desired loss function $\mathcal{L}$, and $\wp_C$ is the projection operator on C . The projection operator $\wp_C$ locates a point in C having the least distance to a given point, while v_0, v_1 are initialization points.

AMLA converges to a solution of the deconvolution problem (Equation (2)) for various loss functions, including those reviewed in Section 2.2. A detailed mathematical analysis establishing this convergence is presented in Appendix A, starting with the preliminary mathematical tools, as well as the lemmas, theorems, and their proofs.

For our choice of loss function (Equation (11)), the gradient A is the vector defined by

$$A_j = \frac{2\theta}{M} \sum_{i=1}^M S_{ij} \left(\left(\delta + \sum_{k=1}^N S_{ik} p_k \right)^\theta - b_i \right) \left(\delta + \sum_{k=1}^N S_{ik} p_k \right)^{\theta-1}. \quad (13)$$

Furthermore, the control parameters α_j and λ are determined by a Lipschitz constant (K) of A (See Theorem 12 in Appendix A). For A defined by Equation (13), a suitable K can be given by

$$K = \begin{cases} N \left[\frac{2\theta^2 s_{\max}^2}{\delta^{2(1-\theta)}} + \frac{2\theta|\theta-1|s_{\max}^2[(Ns_{\max}+\delta)^\theta + b_{\max}]}{\delta^{2-\theta}} \right], & \text{for } \theta \leq 1 \\ N \left[2\theta^2 s_{\max}^2 (Ns_{\max} + \delta)^{2(\theta-1)} + \frac{2\theta|\theta-1|s_{\max}^2[(Ns_{\max}+\delta)^\theta + b_{\max}]}{\delta^{2-\theta}} \right], & \text{for } 1 < \theta \leq 2 \\ N \left[2\theta^2 s_{\max}^2 (Ns_{\max} + \delta)^{2(\theta-1)} + 2\theta|\theta-1|s_{\max}^2[(Ns_{\max}+\delta)^\theta + b_{\max}](Ns_{\max} + \delta)^{\theta-2} \right], & \text{for } \theta > 2 \end{cases}, \quad (14)$$

where $s_{\max} = \max_{1 \leq i \leq M} |S_{ij}|$, $b_{\max} = \max_{1 \leq i \leq M} b_i$.

Lastly, the projection operator on C denoted by $\wp_C$ is computed using the alternating projections method [45].

2.5. Validation Datasets

Our validation datasets, which come from published findings in [22,23], consisted of experimentally measured immune cell-type proportions from tumor samples, bulk RNA sequencing data of tumor samples, and a gene expression reference profile. The gene expression reference profile was a signature matrix of eight Melanoma subsets derived from scRNA-seq (SMART-Seq2) (see Supplementary Table S2e in [22]). These melanoma subsets are listed across 3121 genes. They include five TIICs (B cells, CD8 T cells, CD4 T cells, NK cells, and macrophages) and other cell subsets, including endothelial cells, malignant cells, and cancer-associated fibroblasts (CAFs).

Experimentally measured immune cell-type proportions from tumor samples and bulk RNA sequencing data of tumor samples were obtained from [23]. In [23], the authors divided the single-cell suspensions collected from the lymph nodes of four patients with metastatic melanoma into two portions. For one portion, flow cytometry was used to measure the percentages of live cells, including four TIICs (B cells, CD4 T cells, CD8 T cells, and NK cells), malignant cells, and other cells made up of primarily stromal and endothelial cells (see Supplementary Table S3A in [23]). The fractions of each of these cell types for each patient are presented in Table 1 below. We refer to these cell-type fractions as the ground-truth values from the experiment (GTVEs).

The other portion of the single-cell suspensions was used for bulk RNA sequencing (RNA-seq). We downloaded this RNA-seq data for the four Melanoma patients from the “example data from EPIC” link on the EPIC web application (<http://epic.gfellerlab.org>) accessed on 28 September 2022. It can also be accessed from the Gene Expression Omnibus (GEO) repository [46] through the accession number GSE93722. This RNA-seq data consists of 49,902 genes quantified in transcripts per million (TPM) for each of the four melanoma samples.

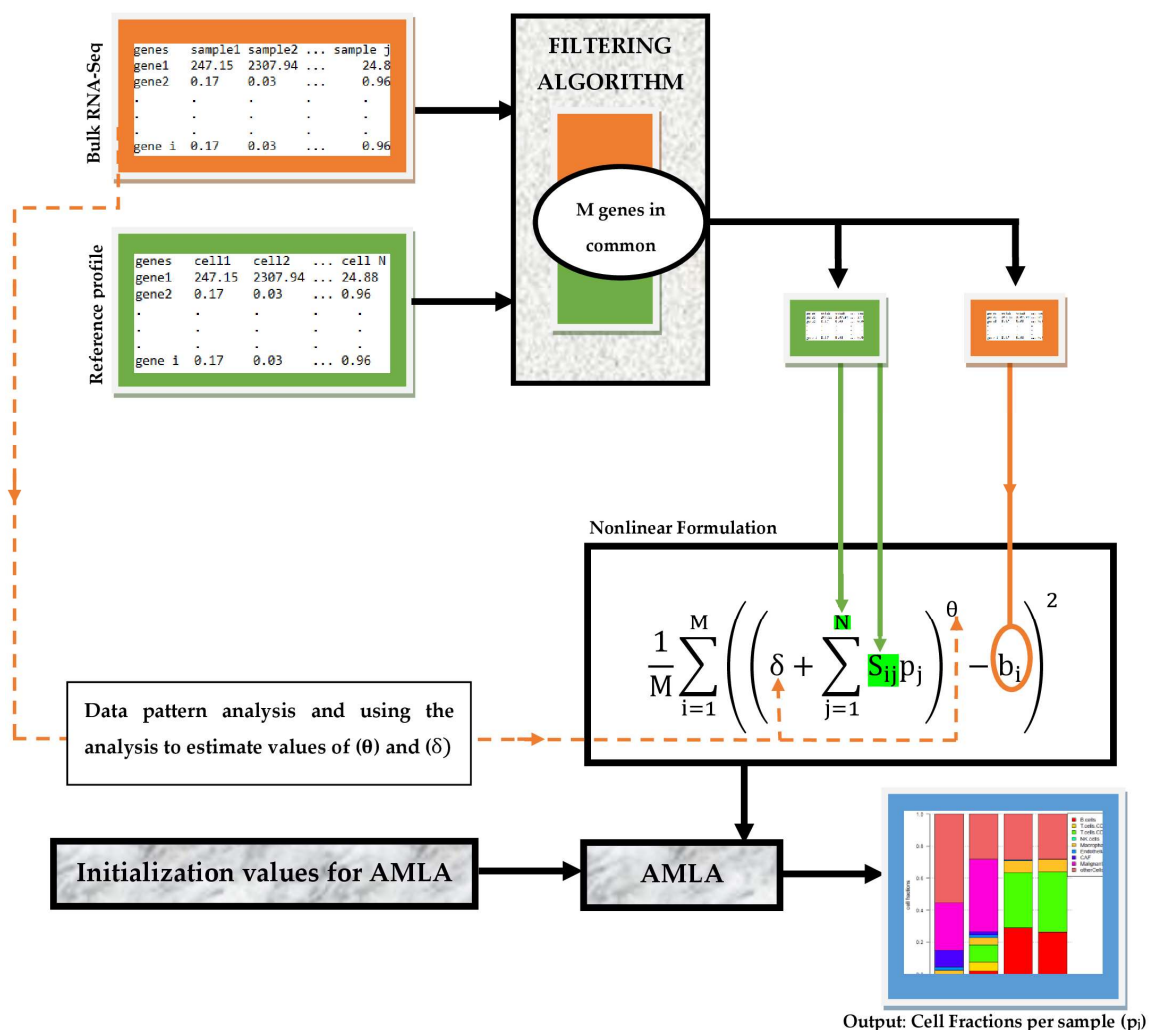
Table 1. Fractions of cell types measured using flow cytometry for the lymph nodes of metastatic melanoma patients (modified from Supplementary Table S3A of [23]).

Patient ID	B Cells	CD4 T Cells	CD8 T Cells	NK Cells	Malignant Cells	Other Cells ¹
LAU125	0.1812	0.0082	0.0035	0.0050	0.6803	0.1218
LAU355	0.3248	0.2315	0.0582	0.0017	0.0006	0.3832
LAU1255	0.0579	0.0276	0.0376	0.0017	0.3756	0.4997
LAU1314	0.4667	0.1815	0.0454	0.0025	0.0007	0.3031

¹ These consist mostly of stromal (for example, cancer-associated fibroblasts (CAFs)) and endothelial cells.

2.6. Deconvolution Workflow

Our deconvolution workflow consists of partly sequential steps necessary to achieve efficient and accurate estimation of TIICs from bulk RNA-seq data. As illustrated in Figure 1, the input data comprise a tab-delimited text file of bulk RNA-seq samples and gene reference profiles. These inputs are first processed using a simple data filter algorithm, which identifies the genes common to both inputs and passes values of these genes in each input data across the filter. These values are then fed into the respective variables of our nonlinear framework. After that, the cell fractions per sample are computed using AMLA. We remark that suitable values of the parameters θ and δ can be estimated using a rigorous non-trivial pattern analysis of the bulk RNA-Seq data, as indicated in Figure 1, although no attempt was made in this direction in the present work.

**Figure 1.** Workflow for the deconvolution of TIICs using a nonlinear optimization framework.

2.7. Software Used

We implement the deconvolution workflow described in Figure 1 as a Python package. The package was developed using the IDE (Integrated Development Environment) PyCharm Community Edition 2021.2.1 version 212.5080.64 created by JetBrains s.r.o, Prague, Czech Republic, running Python 3.9.7 (64 bit) version 3.9.7150.0. The package contains three custom-made modules: the filtering algorithm, our nonlinear framework, and AMLA. The Pandas library was used to manipulate the import of sequencing data and reference profiles and export of estimated fractions for visualization.

To compare our method with two popular cell-type deconvolution methods, we also generated results from CIBERSORTx [22] and EPIC [23] using the web application versions of their software available at <https://cibersortx.stanford.edu/> accessed on 30 September 2022 and <http://epic.gfellerlab.org> accessed on 28 September 2022, respectively. We performed all these software activities on an Intel® Core™ i5-6300U CPU @at 2.40GHz with 8 GB RAM on a 64 bit Windows 10 Pro operating system.

3. Results

3.1. Estimating Cell-Type Fractions in Four Melanoma Samples Using Our Nonlinear Framework, EPIC, and CIBERSORTx

We considered the bulk RNA-Seq dataset of four melanoma patients and the gene reference profile containing eight cell subsets, as described in Section 2.5. By inputting tab-delimited text files of these two datasets into the filtering algorithm, we obtained 2928 genes common to both datasets. The values for these specific genes in the respective datasets were passed across the filtering algorithm into the nonlinear framework. We estimated eight cell-type fractions present in each of the four melanoma samples using four different versions of our nonlinear framework. These versions, named according to the value of the hyperparameter θ in Equation (11) and the procedure for applying AMLA, include those described below.

1. Equivalent linear model (ELM), expressed for $\theta = 1$, such that Equation (11) becomes

$$\mathcal{L}(B, r(S, P)) = \frac{1}{M} \sum_{i=1}^M \left(\left(\delta + \sum_{j=1}^N S_{ij} p_j \right) - b_i \right)^2. \quad (15)$$

AMLA is then applied to approximate the solution.

2. Linearized nonlinear model (LNM), expressed for $\theta = 0.92$, such that Equation (11) becomes

$$\mathcal{L}(B, r(S, P)) = \frac{1}{M} \sum_{i=1}^M \left(\left(\delta + \sum_{j=1}^N S_{ij} p_j \right)^{0.92} - b_i \right)^2. \quad (16)$$

However, we do not apply AMLA directly to Equation (16). Rather, we linearize it to obtain the form

$$\mathcal{L}(B, r(S, P)) = \frac{1}{M} \sum_{i=1}^M \left(\left(\delta + \sum_{j=1}^N S_{ij} p_j \right) - (b_i)^{\frac{1}{0.92}} \right)^2, \quad (17)$$

and thereafter apply AMLA to approximate the solution.

3. Nonlinear model one (NM1), expressed for $\theta = 0.92$, such that we obtain Equation (16) above, and then apply AMLA to approximate the solution.
4. Nonlinear model two (NM2), expressed for $\theta = 1.08$, such that Equation (11) becomes

$$\mathcal{L}(B, r(S, P)) = \frac{1}{M} \sum_{i=1}^M \left(\left(\delta + \sum_{j=1}^N S_{ij} p_j \right)^{1.08} - b_i \right)^2. \quad (18)$$

AMLA is then applied to approximate the solution.

For all four versions enumerated above, we set the variable δ , such that $\delta = 1$. Moreover, we initialized AMLA with distinct set values of v_0 and v_1 , creating a cartesian product for the patient series. Table 2 summarizes the hyperparameter values δ and θ chosen for the named versions of our nonlinear framework.

Table 2. Hyperparameter values for named versions of our nonlinear framework (Equation (11)).

Version	θ	δ
ELM	1.00	1.00
LNLM	0.92	1.00
NM1	0.92	1.00
NM2	1.08	1.00

We have already emphasized that the nonlinearity of our model (Equation (11)) strictly depends on the value of θ , which must be different from one. Furthermore, we remarked that the selection and tuning procedure for this parameter can be achieved analytically on the basis of the input gene sequencing datasets using rigorous data assimilation techniques. However, our primary goal in this work was to demonstrate that a nonlinear regression approach for the TIIC deconvolution problem could yield significantly more accurate estimates of the fractions of cell types, including TIICs from the bulk gene expression data of tumor samples. Thus, we favored an empirical approach for selecting the hyperparameter θ , in line with this goal.

Our empirical approach relies on the observation that the relationship between the gene expression profiles of pure and heterogeneous samples is not strictly linear. Guided by this, we interpret the expression “not strictly linear” as slightly different from one, and this difference can be either side of one, i.e., greater than or less than one. Because θ is strictly positive, we looked at $\theta \in (0, 1) \cup (1, +\infty)$. Guided by our hypothesis presented in Section 2.1, we randomly selected 0.92 from the interval $(0, 1)$. We also selected 1.08 from the interval $(1, +\infty)$ by considering the symmetric distance of the previous selection from one. The hyperparameter δ is a smoothing parameter. The literature is filled with several rigorous techniques for smoothing parameter estimation. Again, for the same reasons as in θ , we chose to assume a default value of one.

We present two results when estimating the cell-type fractions in the four melanoma samples using EPIC. The first result which we denote as “EPIC1” was obtained from the default setting of the EPIC web application (<http://epic.gfellerlab.org>) accessed on 30 September 2022. The default setting comprises tab-delimited inputs of bulk RNA-seq dataset as described in Section 2.5. Furthermore, it includes a reference profile of seven cell subsets (B cells, CAFs, CD4 T cells, CD8 T cells, endothelial, macrophages, and NK cells), built from tumor-infiltrating cells from TPM normalized scRNA-seq (see Supplementary Table S2A in [23]). This reference profile contains 23,684 genes. The second result, which we denote as “EPIC2”, was obtained using the datasets as described in Section 2.5.

Furthermore, we estimated the cell-type fractions in the four melanoma samples using the CIBERSORTx web application (<https://cibersortx.stanford.edu/>) accessed on 30 September 2022. The gene expression reference profile used was as described in Section 2.5. Here, we present two results for our cell-type fractions estimation using CIBERSORTx. The first result, which we denote as “CIBERSORTx1”, was obtained by checking the batch correction box, selecting B-mode, and then running CIBERSORTx, after the tab-delimited bulk RNA-seq and reference profile files were uploaded. The second result, which we denote as “CIBERSORTx2”, was obtained following the same procedure as in the first result, with the only exception being to uncheck the batch correction box. In both estimations using CIBERSORTx, quantile normalization was disabled, and the permutation for significance analysis was set at 100. We present these results in Table 3.

Table 3. Fractions of cell types estimated using deconvolution methods for the lymph nodes of metastatic melanoma patients.

Patient ID	Deconvolution Method	B Cells	CD8 T Cells	CD4 T Cells	NK Cells	Macrophages	Endothelial Cells	CAF	Malignant Cells
LAU125	ELM	0.0139	0.0450	0.0714	0.0376	0.0389	0.0628	0.1175	0.6129
	LNM	0.0009	0.0170	0.0283	0.0120	0.0872	0.0468	0.1165	0.6834
	NM1	0.1884	0.0063	0.0000	0.0158	0.0684	0.0000	0.0616	0.6594
	NM2	0.1884	0.0063	0.0000	0.0158	0.0684	0.0000	0.0616	0.6594
	EPIC1 *	0.0101	0.0095	0.0303	0.0000	0.0120	0.0253	0.0003	-
	EPIC2 **	0.0000	0.0000	0.0000	0.0000	0.0258	0.0161	0.1087	0.2987
	CIBERSORTx1	0.0015	0.0000	0.0250	0.0000	0.0175	0.0002	0.0063	0.9494
	CIBERSORTx2	0.0000	0.0000	0.0107	0.0000	0.0120	0.0000	0.0135	0.9638
LAU355	ELM	0.1596	0.1154	0.3119	0.0264	0.0865	0.1008	0.0970	0.1023
	LNM	0.2558	0.1085	0.2828	0.0049	0.2287	0.0680	0.0331	0.0182
	NM1	0.3221	0.0836	0.2299	0.0084	0.0821	0.2122	0.0616	0.0000
	NM2	0.3220	0.0837	0.2300	0.0084	0.0821	0.2122	0.0616	0.0000
	EPIC1 *	0.4540	0.0182	0.2672	0.0000	0.0086	0.0000	0.0001	-
	EPIC2 **	0.1834	0.0000	0.4528	0.0000	0.1058	0.0034	0.0000	0.0000
	CIBERSORTx1	0.5550	0.0000	0.3536	0.0104	0.0794	0.0000	0.0000	0.0017
	CIBERSORTx2	0.5896	0.0000	0.3297	0.0065	0.0741	0.0000	0.0000	0.0000
LAU1255	ELM	0.0383	0.1102	0.1270	0.0390	0.0510	0.0981	0.1070	0.4292
	LNM	0.0346	0.0981	0.0985	0.0108	0.1111	0.0429	0.0610	0.5431
	NM1	0.0589	0.0521	0.0342	0.0068	0.1042	0.2974	0.1089	0.3373
	NM2	0.0589	0.0521	0.0342	0.0069	0.1042	0.2974	0.1089	0.3374
	EPIC1 *	0.0411	0.1299	0.0583	0.0000	0.0197	0.0000	0.0001	-
	EPIC2 **	0.0148	0.0563	0.1094	0.0000	0.0487	0.0138	0.0216	0.4628
	CIBERSORTx1	0.0493	0.0987	0.1059	0.0000	0.1360	0.0003	0.0035	0.6062
	CIBERSORTx2	0.0329	0.0993	0.0707	0.0000	0.1490	0.0002	0.0018	0.6462
LAU1314	ELM	0.1773	0.1040	0.2823	0.0343	0.0898	0.0890	0.0978	0.1257
	LNM	0.2872	0.1026	0.2519	0.0089	0.2433	0.0515	0.0284	0.0261
	NM1	0.4436	0.0452	0.1695	0.0032	0.0095	0.2506	0.0658	0.0126
	NM2	0.4436	0.0453	0.1695	0.0032	0.0094	0.2506	0.0658	0.0126
	EPIC1 *	0.6760	0.0181	0.0790	0.0042	0.0015	0.0000	0.0000	-
	EPIC2 **	0.2040	0.0001	0.4244	0.0000	0.1057	0.0045	0.0000	0.0000
	CIBERSORTx1	0.6183	0.0000	0.3229	0.0062	0.0207	0.0000	0.0000	0.0318
	CIBERSORTx2	0.6593	0.0000	0.3082	0.0099	0.0109	0.0000	0.0000	0.0115

* The reference profile does not contain malignant cells. A column labeled as “other cells” is included in the results with the values 0.9127, 0.2519, 0.7510, and 0.2212 recorded for LAU125, LAU355, LAU1255, and LAU1314, respectively. ** The results also include a column for “other cells” with values 0.5507, 0.2546, 0.2726, and 0.2613 recorded for LAU125, LAU355, LAU1255, and LAU1314, respectively.

It is easy to notice that many values for the deconvolution methods NM1 and NM2 were identical across Table 3. Although NM1 and NM2 had different θ values of 0.92 and 1.08, respectively, both θ values had a symmetric distance of 0.08 about 1.00. This observation directly suggests that the scalar θ (for $\theta \neq 1$) in our nonlinear framework exhibits some symmetry about one, such that we can expect identical results for two choices of θ with the same symmetrical distance about 1.00.

3.2. Estimated vs. Experimentally Measured Cell-Type Fractions in the Four Melanoma Samples

We directly compare the estimated cell-type fractions in the four Melanoma samples presented in Table 3 with the ground-truth values from experiment (GTVE) presented in Table 1. To be able to do this, we ignored the EPIC1 results (due to the absence of malignant fractions) and NM2 results (as they are almost identical to NM1 results). Then, we aggregated the fractions—macrophages, endothelial cells, and CAFs—together as other cells since they satisfied the definition of other cells, as in Table 1. For EPIC2, we also added the values of “other cells” specified in the footnote below Table 3. The comparison is readily visualized in Figure 2.

It is clear from Figure 2 that estimates from NM1 most closely matched the GTVE across all four samples, as indicated by the close resemblance of NM1 and GTVE stacks in all four samples. Even so, stacks from other versions of our nonlinear framework (ELM and LNM) appeared to more closely resemble the GTVE stacks in all four samples than the stacks from EPIC and CIBERSORTx.

However, to quantitatively describe the extent of these resemblances, we considered a general-purpose error metric known as the root-mean-squared error (RMSE). RMSE is excellent for comparing the prediction error of different models for a specific variable. As a result, it is an incredibly good measure of model accuracy. The most accurate model would have an RMSE of zero, which is far from possible. Therefore, the model accuracy is determined by how close the RMSE is to zero, although this determination is made relative to the values of the observations or predictions.

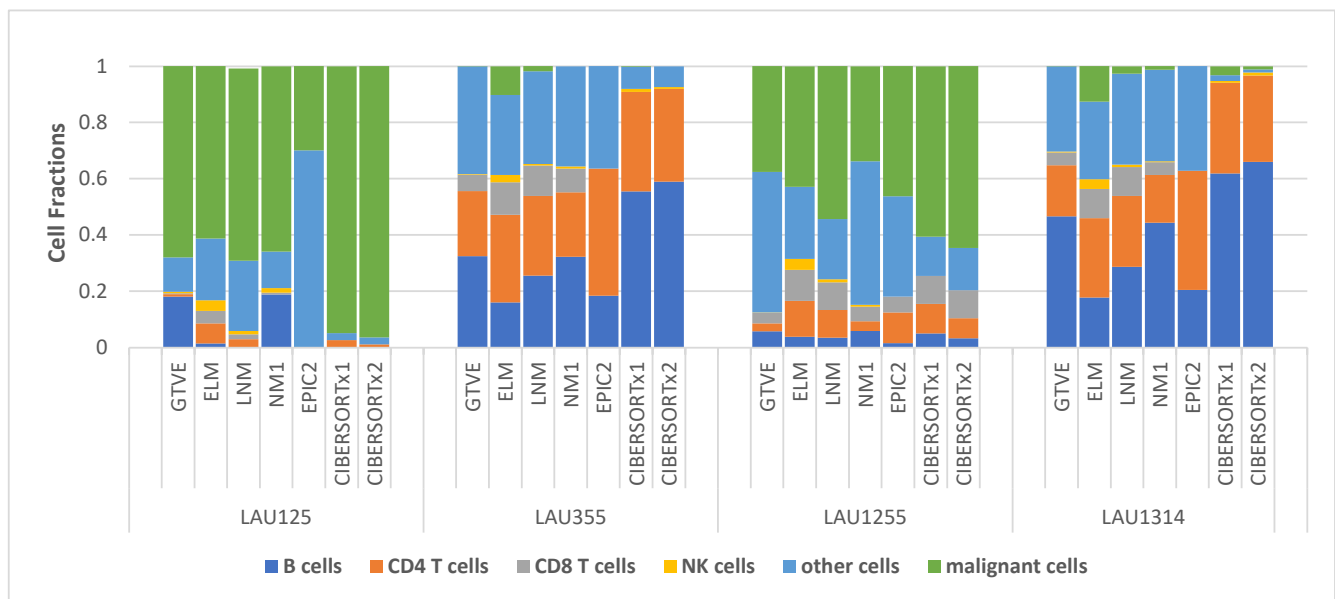


Figure 2. Stacked cell-type fractions per melanoma sample comparing results from different deconvolution methods with ground-truth experimental values.

Here, we calculated the RMSE for NM1, LNM, ELM, CIBERSORTx1, CIBERSORTx2, and EPIC2 for all cell types of the four melanoma samples and then for TIICs only. In the latter, we also included the RMSE calculation for EPIC2. We present these results in Figure 3. We calculated the RMSE values using the formula

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{N}}, \quad (19)$$

where N is the total number of analyzed cell-type subsets, x_i is the GTVE of the analyzed cell-type subset, and $\hat{x}_i$ is the estimated value of the analyzed cell-type subset from a deconvolution method. Values from Tables 1 and 3 were utilized in these calculations.

NM1 can be seen to have the lowest RMSE in both charts of Figure 3. Remarkably, its RMSE value was significantly lower than that of all the other deconvolution methods compared and was extremely close to zero, being in the range of 0–0.02. This indicates that NM1 outperformed the other deconvolution methods and gave more accurate estimates that closely matched GTVE. The accuracy of NM1 can be directly attributed to the nonlinear framework used and the choice of the value of θ . Thus, we validated our hypothesis that a selection of θ slightly greater than or less than one on the order of a few decimal places accurately captures the not strictly linear sense of the relationship between the bulk gene expression and the reference profile. By using the same reference profile in the direct comparison of the deconvolution methods, we successfully limited any chance that the results are a direct consequence of a factor other than the setting or framework of the deconvolution problem.

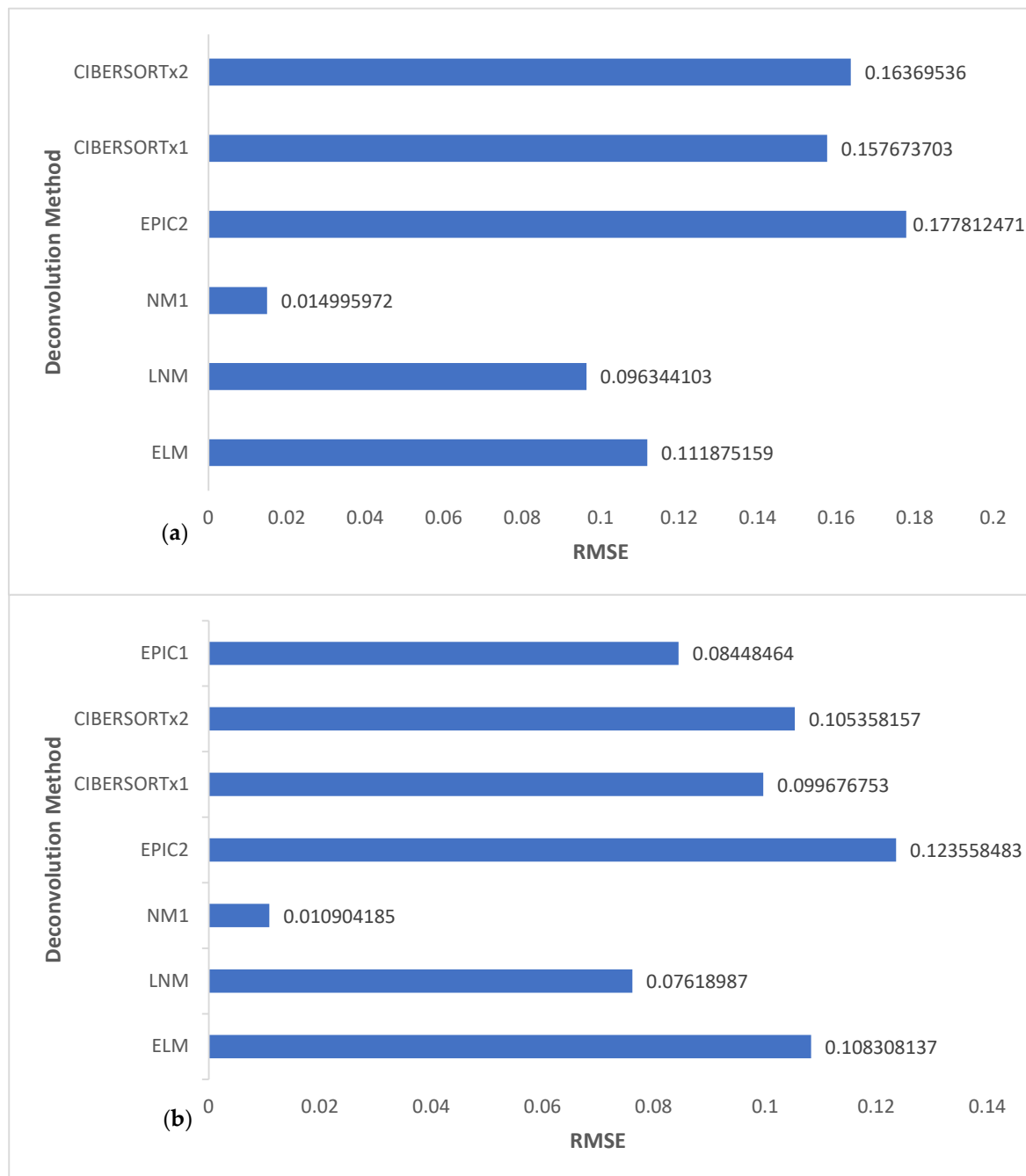


Figure 3. Comparison of the RMSE values of versions of the nonlinear framework, EPIC, and CIBERSORTx for (a) 24 observations including six cell subsets of the four melanoma samples, and (b) 16 observations including only the four TIICs of the four melanoma samples.

4. Discussion

The linear framework has been the dominating paradigm in the computational deconvolution of TIICs and other tumor cell subsets from bulk gene expression data of tumor samples. Efforts toward improving the accuracy of cell fraction estimates from the computational deconvolution of bulk gene expression data have been centered on modifications still built around the basic linear framework. For instance, in both charts of Figure 3, CIBERSORTx1 had a slightly lower RMSE value when compared to CIBERSORTx2, indicating improved accuracy. Similar to the conclusion in [47], we attribute this improved accuracy to the batch correction effect in CIBERSORTx1, which aims to reduce data variability resulting from technical differences between samples. In another instance, in Figure 3b, EPIC1 had a

lower RMSE value when compared to EPIC2, CIBERSORTx1, CIBERSORTx2, and ELM, possibly because of the use of a specific TIIC reference profile. On the other hand, in Figure 3a, ELM, a linear version of our nonlinear framework, had a lower RMSE value (more accuracy) when compared to EPIC2, CIBERSORTx1, and CIBERSORTx2. A likely reason is that the computational algorithm AMLA used in ELM projects directly onto the natural constraint set, which is much different from the computation in CIBERSORTx and EPIC, applied on some broadly defined constraints set, followed by a renormalization of the obtained values.

It is clear from Figure 3 that these modifications would only yield minor improvements in accuracy since they are all based on the linear framework. Notably, the RMSE values from these linear framework-based models revolved around a narrow range of 0.11–0.18 and 0.08–0.11 in Figure 3a,b, respectively. We may not expect any result different from these ranges if we were to analyze these datasets using other deconvolution methods, for example [24–26,31–33], because they are all based on the linear model with variations being in the choice of loss function or other technical modifications.

As shown in Figure 3, the RMSE of NM1 is significantly low, and there is a remarkable difference between its RMSE range and that of the linear models. This observation demonstrates the enormous positive gains in terms of accuracy associated with modeling the deconvolution problem within a nonlinear framework, as it truly represents the natural state of the problem. We also emphasize through LNM results that any attempts to linearize the nonlinear framework before applying the solution algorithm (in this case, AMLA) would vastly diminish these positive gains. However, the outcome may still be slightly better than those from the traditional linear modeling. As evident from Figure 3, LNM RMSE values range from 0.07 to 0.10, which is significantly distant from the NM1 RMSE range (0–0.02) but much closer to the linear models' RMSE ranges. For this reason, AMLA was especially designed to approximate solutions directly from the nonlinear framework without the need to linearize first.

Furthermore, AMLA's design allows it to exhibit faster rates of convergence and runtime in comparison to other traditionally used algorithms in machine learning, which is highly advantageous in the event of enormous amounts of bulk gene expression data of many tumor series. Verifying these with simple numerical experiments on $\mathbb{R}$ (set of real numbers) using known loss functions employed in regression analysis is straightforward. We consider the log-hyperbolic loss, squared error loss, Cauchy loss, and ε -insensitive loss given, for example, by Equations (20)–(23), respectively.

$$q(x) = \log(\cosh(6x - 2)). \quad (20)$$

$$q(x) = (5x - 4)^2. \quad (21)$$

$$q(x) = 5 \log \left[1 + \frac{(3x - 2)^2}{5} \right]. \quad (22)$$

$$q(x) = \begin{cases} -x - \varepsilon, & \text{if } x \in (-\infty, -(1 + \rho)\varepsilon) \\ \frac{(x + (1 - \rho)\varepsilon)^2}{4\rho\varepsilon}, & \text{if } x \in [-(1 + \rho)\varepsilon, -(1 - \rho)\varepsilon] \\ 0, & \text{if } x \in (-(1 - \rho)\varepsilon, (1 - \rho)\varepsilon) \\ \frac{(x - (1 - \rho)\varepsilon)^2}{4\rho\varepsilon}, & \text{if } x \in [(1 - \rho)\varepsilon, (1 + \rho)\varepsilon] \\ x - \varepsilon, & \text{if } x \in ((1 + \rho)\varepsilon, +\infty) \end{cases}. \quad (23)$$

For each of the Equations (20)–(23), we approximated their solutions using AMLA, the well-known classical gradient descent algorithm (CGDA), and Nesterov's accelerated gradient (NAG) widely used in machine learning. Since the optimal solutions of the equations are known, we measured the convergence to the solutions from AMLA, CGDA, and NAG using variations of $(x_n - \bar{x})_{n \in \mathbb{N}}$, where x_n is the labeling obtained at the n -th iteration, and $\bar{x}$ is the optimal solution. We plot this measure of convergence against the number of iterations in Figure 4.

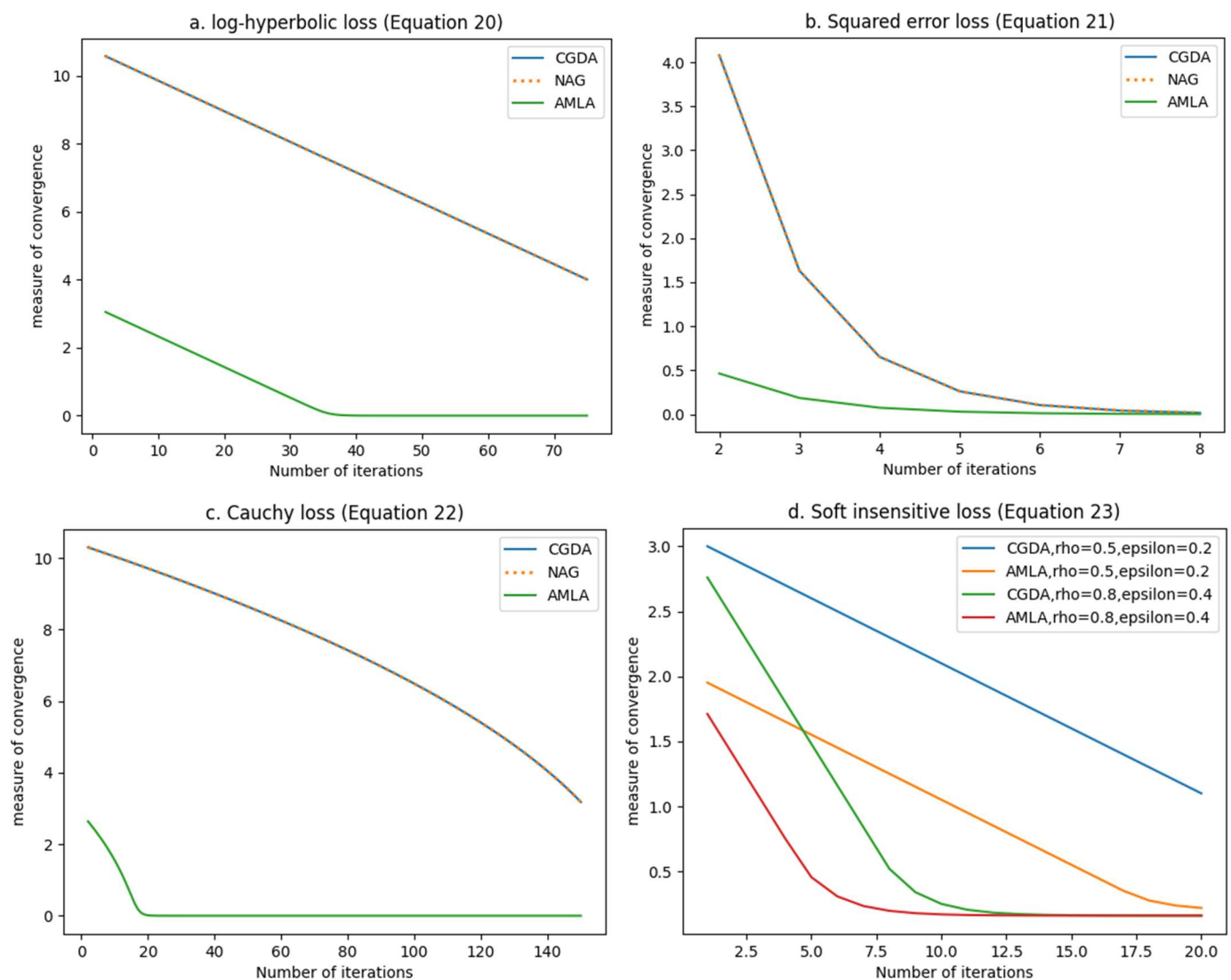


Figure 4. Comparison of the algorithmic efficiency of AMLA, NAG, and CGDA in approximating solutions for (a) log-hyperbolic loss (Equation (20)), (b) squared error loss (Equation (21)), (c) Cauchy loss (Equation (22)), and (d) ϵ -insensitive loss (Equation (23)).

From Figure 4, we can observe that AMLA converged in a significantly fewer number of iterations than CGDA and NAG, for all four loss functions considered. The implication of this observation is that AMLA has a higher order of convergence and, consequently, a faster rate of convergence since the “order of convergence defines the rate of convergence” [48]. Furthermore, Figure 4 affirms the robustness of AMLA. A robust algorithm is one that is theoretically guaranteed to converge, “starting from any initial design estimate” [49], “such that their correctness is not destroyed by round-off errors” [50]. We show in Appendix A, using rigorous mathematical analysis, that AMLA is guaranteed to converge to a solution of the constrained optimization problem (Equation (2)) for a variety of loss functions satisfying the stated conditions. The four loss functions presented in Figure 4 satisfy the stated conditions, thus leading to their convergence in Figure 4, even when CGDA and NAG did not converge (see Figure 4a,c). Moreover, as shown in Figure 4a–d, once AMLA converged, it maintained the flat zero line, which highlights its gross insensitivity to round-off errors, thus affirming its high robustness.

A very noteworthy affirmation of the high robustness of AMLA can be seen in Figure 4d. The ϵ -insensitive loss introduces extra parameters of ρ and ϵ whose selection can critically affect the convergence and robustness of any machine learning algorithm. As

shown in Figure 4d, when we randomly selected $\rho = 0.5$ and $\varepsilon = 0.2$, AMLA converged in fewer than 20 iterations while CGDA was yet to converge beyond 20 iterations. Furthermore, when we randomly selected $\rho = 0.8$ and $\varepsilon = 0.4$, AMLA converged in about seven iterations, while CGDA converged in about nine iterations. These observations from Figure 4d clearly indicate that the values of the parameters ρ and ε are significant determinants of the rate of convergence of AMLA. In fact, from Figure 4d, we can see that the selection of ρ and ε can either increase or decrease the rate of convergence of AMLA. However, it does not affect the robustness of AMLA, as AMLA is guaranteed to converge regardless of the values of ρ and ε . It is important to point out that, as seen in Figure 4d, AMLA still outperformed CGDA for the chosen values of ρ and ε .

Overall, this proof of principle demonstrates that fully accurate and efficient computational deconvolution of tumor bulk gene expression data for estimation of the proportions of cell-type fractions, including TIICs, is best achievable using a nonlinear optimization framework, whose solution can be approximated by an accelerated machine learning algorithm (AMLA). This nonlinear optimization framework truly captures the natural state of the deconvolution problem. Consequently, in the future, we will implement the entire deconvolution workflow, described in Figure 1, as a cloud-based tool with a user-friendly graphical interface, which we shall call NECSTGEP (Naturally Estimating Cell-type Subsets from Tumor Gene Expression Profiles). NECSTGEP will be equipped with an additional rigorous machine learning algorithm that will be able to automatically fix the model parameters and the initialization values for AMLA, using pattern analysis of the input bulk gene expression profiles, as well as the reference profile. This is very essential as this proof of principle has shown a crucial role in yielding highly accurate estimation results. Similar problems in biology, oncology, and medical physics, such as the optimal scheduling of combined cancer therapies and reconstruction of gene regulatory networks, parade similar levels of complexity. Thus, they can benefit from an application of the technique described thus far, to yield highly accurate results.

5. Conclusions

We introduced and discussed a nonlinear constrained optimization framework for the computational deconvolution of TIICs and other tumor cell-type subsets from tumor bulk gene expression profiles, in addition to an accelerated machine learning algorithm (AMLA) for directly approximating its solution. Our analysis using real tumor transcriptomics datasets concluded that this nonlinear approach yields values closely matching ground-truth values from experiment, because it treats the problem in its natural state. Models NM1 and NM2 produced the “best” values for the estimated cell-type fractions in this study and were significantly different from those obtained using the traditional linear modeling approach. However, one main limitation of the study is the empirical choice of model hyperparameters, which will be addressed in future studies. This study, therefore, heralds a paradigm shift away from the traditional linear modeling of the TIIC deconvolution problem.

Author Contributions: Conceptualization, L.C.O. and E.A.O.; methodology, L.C.O., A.U.B. and E.A.O.; software, L.C.O. and E.A.O.; validation, L.C.O. and A.U.B.; formal analysis, L.C.O. and A.U.B.; investigation, L.C.O., A.U.B. and E.A.O.; resources, L.C.O. and A.U.B.; data curation, L.C.O. and E.A.O.; writing—original draft preparation, L.C.O. and E.A.O.; writing—review and editing, L.C.O., A.U.B. and E.A.O.; visualization, L.C.O. and E.A.O.; supervision, A.U.B.; project administration, L.C.O. and A.U.B.; funding acquisition, L.C.O. and A.U.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the UK government through the Commonwealth Scholarship, grant number NGCN-2020-263. L.C.O. acknowledges financial endowment from Foundation L’Oreal and UNESCO through the 2021 L’Oreal–UNESCO For Women in Science Young Talent Award Sub-Saharan Africa.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. The data can be found at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93722>. The data presented in this study are openly available from <https://doi.org/10.7554/eLife.26476.023>, <https://doi.org/10.7554/eLife.26476.024>, <https://doi.org/10.1038/s41587-019-0114-2>.

Acknowledgement: The authors are grateful for the comments of the two anonymous reviewers which helped to improve this work.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Let C be a nonempty, closed, and convex subset of a Hilbert space $\mathbb{H}$.

Definition 1. A vector $v := \wp_C(u) \in C$ is called the projection of $u \in \mathbb{H}$ onto C if and only if $v = \inf_{z \in C} \|u - z\|^2$. Equivalently, if and only if $\langle v - u, v - z \rangle \leq 0 \forall z \in C$.

Definition 2. A mapping Γ (a) is said to be monotone on C if and only if $\langle \Gamma u - \Gamma v, u - v \rangle \geq 0 \forall (u, v) \in C \times C$. Furthermore, if there exists $\xi > 0$ such that $\langle \Gamma u - \Gamma v, u - v \rangle \geq \xi \|u - v\|^2$, $\forall (u, v) \in C \times C$, then Γ is called ξ -strongly monotone. (b) is said to be Lipschitz on C , if there exists a scalar $K > 0$ such that $\|\Gamma u - \Gamma v\| \leq K \|u - v\|$, $\forall (u, v) \in C \times C$. A monotone operator is said to be maximal monotone if it has no proper monotone extension.

Definition 3. A subset $\mathbb{S}$ of $\mathbb{H}$ is said to be bounded if and only if there exists a scalar $\mu > 0$ such that $\|u\| \leq \mu \forall u \in \mathbb{S}$.

Definition 4. Let g be a differentiable mapping on C . A vector $u^* \in C$ is called a stationary (critical) point of g if $\langle \nabla_g(u^*), u - u^* \rangle \geq 0 \forall u \in C$.

Definition 5. Let $g: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous mapping and ∂g denote the subdifferential of g . g is said to have the Kurdyka–Lojasiewicz (KL) property at $u^* \in \text{dom} \partial g = \{u \in \mathbb{R}^n : \partial g(u) \neq \emptyset\}$, if there exists $\eta \in (0, +\infty]$, a neighborhood V of u^* , and a continuous concave map $\varphi: [0, \eta] \rightarrow \mathbb{R}^+$ with $\varphi(0) = 0$, φ is $\mathcal{C}^1$ on $(0, \eta)$, $\varphi'(s) > 0 \forall s \in (0, \eta)$ and $\varphi'(g(u) - g(u^*)) \text{dist}(0, \partial g(u)) \geq 1 \forall u \in V \cap \{v \in \mathbb{R}^n : g(u^*) < g(v) < g(u^*) + \eta\}$, where $\text{dist}(z, \mathbb{S}) = \inf_{s' \in \mathbb{S}} \|z - s'\|^2$. Any g satisfying the KL property at each point of $\text{dom} \partial g$, is called the KL function. The set of KL functions is rich, including as subsets, real and sub-analytic functions, real semi-algebraic functions, semi-convex functions, uniformly convex functions, and convex functions satisfying a growth condition (see [51–53] and references contained therein).

Lemma 6. Let g be a differentiable mapping on C with K -Lipschitz gradient ∇_g ; then, $|g(u) - g(v) - \langle u - v, \nabla_g(v) \rangle| \leq \frac{K}{2} \|u - v\|^2 \forall u, v \in C$.

Proof. Let g_0 be a real valued map on $\mathbb{R}$ defined by $g_0(t) = g(v + t(u - v))$; then, it follows that $g_0(0) = g(v)$, $g_0(1) = g(u)$ and

$$g(u) - g(v) = g_0(1) - g_0(0) = \int_0^1 g_0'(t) dt.$$

Now, $g_0'(t) = \langle u - v, \nabla_g(v + t(u - v)) \rangle$; thus,

$$\begin{aligned}
& |g(u) - g(v) - \langle u - v, \nabla_g(v) \rangle| \\
&= \left| \int_0^1 \langle u - v, \nabla_g(v + t(u - v)) \rangle dt - \langle u - v, \nabla_g(v) \rangle \right| \\
&\leq \int_0^1 |\langle u - v, \nabla_g(v + t(u - v)) - \nabla_g(v) \rangle| dt \\
&\leq \int_0^1 \|\nabla_g(v + t(u - v)) - \nabla_g(v)\| \|u - v\| dt \\
&\leq \int_0^1 Kt \|u - v\|^2 dt = \frac{K}{2} \|u - v\|^2. \square
\end{aligned}$$

Lemma 7. [54] Let $\{u_n\}_{n \geq 0} \subseteq \mathbb{H}$ be a bounded sequence in $\mathbb{H}$. Then, there exists a subsequence $\{u_{n_j}\}_{j \geq 0} \subseteq \{u_n\}_{n \geq 0}$ that converges weakly (denoted as $u_{n_j} \rightharpoonup u_*$ (say)) in $\mathbb{H}$. Note that, for $\mathbb{H} = \mathbb{R}^N$, strong and weak convergence coincide.

Lemma 8. Let $\{u_n\}_{n \geq 0}$ and $\{z_n\}_{n \geq 0}$ be real sequences such that $u_n \geq 0 \forall n \geq 0$, $\lim_{n \rightarrow \infty} \sum_{i=0}^n z_i \in \mathbb{R}$ and $u_{n+1} \leq u_n + z_n \forall n \geq 0$. Then $\lim_{n \rightarrow \infty} u_n$ exists in $\mathbb{R}$.

Proof. We follow a method of proof in [55]. Let, $\lim_{n \rightarrow \infty} \sum_{i=0}^n z_i = z_*$.

Define $y_n = \sum_{i=0}^{n-1} z_i \forall n \geq 1$; then, $\lim_{n \rightarrow \infty} y_n = z_*$.

Now, $u_{n+1} + y_n \leq u_n + z_n + y_n = u_n + y_{n+1} \forall n \geq 1$.

Thus, $u_{n+1} - y_{n+1} \leq u_n - y_n \forall n \geq 1$. Therefore, the sequence $\{u_n - y_n\}_{n \geq 1}$ is monotone nonincreasing. Hence, only the following two cases are possible:

Case 1: $\lim_{n \rightarrow \infty} (u_n - y_n) = -\infty$,

Case 2: $\lim_{n \rightarrow \infty} (u_n - y_n) \in \mathbb{R}$.

It is impossible for case 1 to be true because of the following contradiction:

Assuming case 1 holds and using the hypothesis that $u_n \geq 0 \forall n \geq 1$,

$$0 \leq \lim_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} ((u_n - y_n) + y_n) = -\infty.$$

Hence, case 2 must be true. Consequently, we have that

$$\lim_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} ((u_n - y_n) + y_n) = \lim_{n \rightarrow \infty} (u_n - y_n) + \lim_{n \rightarrow \infty} y_n \in \mathbb{R}. \square$$

Lemma 9. [56] Let Γ be a single valued monotone operator on $\mathbb{H}$ such that $C \subseteq \text{dom}\Gamma = \{u \in \mathbb{H} : \Gamma(u) \neq \emptyset\}$ and Γ is hemicontinuous on C . Let T_C be the normality operator for C , i.e., $T_C(u) = \{y \in \mathbb{H} : \langle u - v, y \rangle \geq 0, \forall v \in C\}$. Then, $\Gamma + T_C$ is a maximal monotone operator.

Lemma 10. [57] Let $\{u_n\}_{n \geq 0}$ be a sequence in $\mathbb{H}$ that converges weakly to $u_0 \in \mathbb{H}$. Then, for any $u \neq u_0$, $\liminf_{n \rightarrow 0} \|u_n - u_0\| < \liminf_{n \rightarrow 0} \|u_n - u\|$.

Lemma 11. [58] Let $g : \mathbb{R}^{2N} \rightarrow \mathbb{R} \cup \{+\infty\}$ (sic) be a proper lower semicontinuous mapping and $z_n = (u_n, u_{n-1})_{n \geq 1}$ be a sequence satisfying the following:

(H1) for each $n \geq 1$, $g(z_{n+1}) + a\|u_n - u_{n-1}\|^2 \leq g(z_n)$ for some fixed positive constant a ;

(H2) for each $n \geq 1$, there exists $y_{n+1} \in \partial g(z_{n+1})$ such that $\|y_{n+1}\| \leq \frac{b}{2}(\|u_n - u_{n-1}\| + \|u_{n+1} - u_n\|)$ for some fixed positive constant b ;

(H3) there exists a subsequence $(z_{n_k})_{k \geq 1}$ such that $z_{n_k} \rightarrow \hat{z}$ and $g(z_{n_k}) \rightarrow g(\hat{z})$ as $k \rightarrow \infty$. Moreover, let g have the KL property at the cluster point $\hat{z}$ specified in (H3). Then the sequence $\{u_n\}_{n \geq 0}$ has finite length (that is $\sum_{n=1}^{\infty} \|u_n - u_{n-1}\| < +\infty$) and converges to $\hat{u}$ as $n \rightarrow \infty$, where $(\hat{u}, \hat{u})$ is a critical point of g .

Theorem 12. Let g be a real valued mapping on C , bounded below, and whose Frechet derivative denoted by A is K -Lipschitz. Let $\{v_j\}_{j \geq 0}$ be the sequence generated iteratively by AMLA (see Equation (12)), where the control parameters α_j and λ are chosen in $\mathbb{R}$ such that

$$0 \leq \alpha_j < \frac{2-\lambda K}{3}, \lim_{j \rightarrow \infty} \alpha_j = 0, \alpha_{j+1} \leq \alpha_j \forall j \geq 0 \text{ and } 0 < \lambda < \frac{1}{K}. \text{ Then,}$$

- (a) there exists $\mu, \alpha > 0$ such that for the sequence $\{f(v_j)\}_{j \geq 0} = \{g(v_j) + \mu\|v_j - v_{j-1}\|^2\}_{j \geq 0}$, $f(v_{j+1}) + \alpha\|v_{j+1} - v_j\|^2 \leq f(v_j) \forall j \geq 0$;
- (b) $\sum_{j=0}^{\infty} \|v_{j+1} - v_j\|^2 < +\infty$;
- (c) $\lim_{j \rightarrow \infty} g(v_j) \in \mathbb{R}$.

Proof. According to Lemma 6, we have that $g(v_{j+1}) - g(v_j) \leq v_{j+1} - v_j, Av_j + \frac{K}{2}v_{j+1} - v_j^2$. This implies that

$$\begin{aligned} \lambda(g(v_{j+1}) - g(v_j)) &\leq \langle v_{j+1} - v_j, \lambda Av_j \rangle + \frac{\lambda K}{2} \|v_{j+1} - v_j\|^2 \\ &= \langle v_{j+1} - v_j, w_j - (w_j - \lambda Av_j) \rangle + \frac{\lambda K}{2} \|v_{j+1} - v_j\|^2 \\ &= \langle v_{j+1} - v_j, v_{j+1} - (w_j - \lambda Av_j) \rangle + \langle v_{j+1} - v_j, w_j - v_{j+1} \rangle + \frac{\lambda K}{2} \|v_{j+1} - v_j\|^2 \\ &\leq -\left(1 - \frac{\lambda K}{2}\right) \|v_{j+1} - v_j\|^2 + \alpha_j \langle v_{j+1} - v_j, v_{j-1} - v_j \rangle \\ &\leq -\left(1 - \frac{\lambda K}{2}\right) \|v_{j+1} - v_j\|^2 + \alpha_j \langle v_{j+1} - v_j, v_{j-1} - v_j \rangle + \frac{\alpha_j}{2} \|v_{j+1} - v_{j-1}\|^2 \\ &= -\left(1 - \frac{\lambda K}{2}\right) \|v_{j+1} - v_j\|^2 + \frac{\alpha_j}{2} \|v_{j+1} - v_j\|^2 + \frac{\alpha_j}{2} \|v_j - v_{j-1}\|^2 \\ &= -\left(1 - \frac{\lambda K + \alpha_j}{2}\right) \|v_{j+1} - v_j\|^2 + \frac{\alpha_j}{2} \|v_j - v_{j-1}\|^2 \\ &\leq -\left(\frac{2-\lambda K}{3}\right) \|v_{j+1} - v_j\|^2 + \left(\frac{2-\lambda K}{6}\right) \|v_j - v_{j-1}\|^2 \end{aligned}$$

Thus, we have that

$$\begin{aligned} \lambda(g(v_{j+1}) - g(v_j)) + \left(\frac{2-\lambda K}{6}\right) (\|v_{j+1} - v_j\|^2 - \|v_j - v_{j-1}\|^2) \\ \leq -\left(\frac{2-\lambda K}{6}\right) \|v_{j+1} - v_j\|^2. \end{aligned}$$

Define the sequence $\{f(v_{j+1})\}_{j \geq 1}$ as

$$f(v_j) = g(v_j) + \left(\frac{2-\lambda K}{6\lambda}\right) \|v_j - v_{j-1}\|^2.$$

Then, $\{f(v_{j+1})\}_{j \geq 1}$ is bounded below, and it follows that

$$f(v_{j+1}) + \alpha \|v_{j+1} - v_j\|^2 \leq f(v_j) \forall j \geq 1 \text{ where } \alpha = \frac{2-\lambda K}{6\lambda}.$$

Thus, the sequence is monotone nonincreasing and bounded below. Hence, $\lim_{j \rightarrow \infty} f(v_j)$ exists. Moreover, we have that

$$\sum_{j=0}^{\infty} \left(\frac{2 - \lambda K}{6\lambda} \right) \|v_{j+1} - v_j\|^2 < +\infty.$$

Therefore,

$$\sum_{j=0}^{\infty} \|v_{j+1} - v_j\|^2 < +\infty, \lim_{j \rightarrow \infty} \|v_{j+1} - v_j\|^2 = 0 \text{ and } \lim_{j \rightarrow \infty} g(v_j) \text{ exists. } \square$$

Theorem 13. Let the assumptions of Theorem 12 hold. Suppose further that A is monotone; denote by S_{Cg} the set of critical points of g . Let $v^* \in S_{Cg}$ and $\{v_j\}_{j \geq 0}$ be the sequence generated by AMLA. Then, the following applies:

- $\{v_j\}_{j \geq 0}$ is bounded;
- $\lim_{j \rightarrow \infty} \|v_{j+1} - w_j\|^2 = 0$;
- $\lim_{j \rightarrow \infty} \|v_j - v^*\|^2$ exists;
- the sequence $\{v_j\}_{j \geq 0}$ converges weakly to a critical point $\hat{v}$ of g ; moreover, if g is convex, then $\hat{v}$ is a minimizer of g

Proof. Let $u_j := w_j - \lambda A v_j$ and $\tau_j := 2\lambda(g(v_j) - g(v_{j+1})) + \lambda K \|v_{j+1} - v_j\|^2$. Now, using the definition of v_j and w_j , Definition 1, the fact that $v^* \in S_{Cg}$, and the monotonicity of A , we get the following estimation:

$$\begin{aligned} \|v_{j+1} - v^*\|^2 &\leq \|u_j - v^*\|^2 - \|v_{j+1} - u_j\|^2 \\ &= \|w_j - \lambda A v_j - v^*\|^2 - \|v_{j+1} - (w_j - \lambda A v_j)\|^2 \\ &= \|w_j - v^*\|^2 - 2\lambda \langle w_j - v^*, A v_j \rangle + \|\lambda A v_j\|^2 \\ &\quad - (\|v_{j+1} - w_j\|^2 + 2\lambda \langle v_{j+1} - w_j, A v_j \rangle + \|\lambda A v_j\|^2) \\ &= \|w_j - v^*\|^2 - 2\lambda \langle v_{j+1} - v^*, A v_j \rangle - \|v_{j+1} - w_j\|^2 \\ &\leq \|w_j - v^*\|^2 - 2\lambda \langle v_{j+1} - v_j, A v_j \rangle - \|v_{j+1} - w_j\|^2 \\ &\leq (1 - \alpha_j) \|v_j - v^*\|^2 + \alpha_j \|v_{j-1} - v^*\|^2 - 2\lambda \langle v_{j+1} - v_j, A v_j \rangle. \end{aligned} \tag{A1}$$

Next, according to Lemma 6,

$$-2\lambda \langle v_{j+1} - v_j, A v_j \rangle \leq 2\lambda(g(v_j) - g(v_{j+1})) + \lambda K \|v_{j+1} - v_j\|^2.$$

Thus,

$$\begin{aligned} \|v_{j+1} - v^*\|^2 &\leq \|v_j - v^*\|^2 - \alpha_j (\|v_j - v^*\|^2 - \|v_{j-1} - v^*\|^2) \\ &\quad + 2\lambda(g(v_j) - g(v_{j+1})) + \lambda K \|v_{j+1} - v_j\|^2 \\ &\leq \|v_{j-1} - v^*\|^2 - \alpha_{j-1} (\|v_{j-1} - v^*\|^2 - \|v_{j-2} - v^*\|^2) + \tau_{j-1} \\ &\quad - \alpha_j (\|v_j - v^*\|^2 - \|v_{j-1} - v^*\|^2) + \tau_j \\ &\quad \vdots \\ &\leq \|v_1 - v^*\|^2 + \sum_{i=1}^j -\alpha_i (\|v_i - v^*\|^2 - \|v_{i-1} - v^*\|^2) + \sum_{i=1}^j \tau_i \\ &\leq \|v_1 - v^*\|^2 + \alpha_1 \|v_0 - v^*\|^2 - \alpha_j \|v_j - v^*\|^2 + \sum_{i=1}^j \tau_i \\ &\leq \|v_1 - v^*\|^2 + \alpha_1 \|v_0 - v^*\|^2 + M, \forall j \geq 1. \end{aligned}$$

Therefore, $\{v_j\}_{j \geq 0}$ is bounded.

Next, $\|v_{j+1} - v^*\|^2 \leq \|v_j - v^*\|^2 - \alpha_j(\|v_j - v^*\|^2 - \|v_{j-1} - v^*\|^2) + \tau_j$, such that,

$$\|v_{j+1} - v^*\|^2 + \alpha_j \|v_j - v^*\|^2 \leq \|v_j - v^*\|^2 + \alpha_{j-1} \|v_{j-1} - v^*\|^2 + \tau_j. \quad (\text{A2})$$

According to Lemma 8, we have that $\lim_{j \rightarrow \infty} (\|v_j - v^*\|^2 + \alpha_{j-1} \|v_{j-1} - v^*\|^2)$ exists.

Moreover, combining Equations (A1) and (A2) and Theorem 12(b) yields $\lim_{j \rightarrow \infty} \|v_{j+1} - w_j\|^2 = 0$ and $\lim_{j \rightarrow \infty} \|v_j - w_j\| = 0$.

Since $\lim_{j \rightarrow \infty} \alpha_j = 0$ and $\{\|v_j - v^*\|^2\}_{j \geq 0}$ is bounded, then $\lim_{j \rightarrow \infty} \alpha_{j-1} \|v_{j-1} - v^*\|^2 = 0$.

Hence, $\lim_{j \rightarrow \infty} \|v_j - v^*\|^2$ exists.

Now, we show that every weak sequential limit of $\{v_j\}_{j \geq 0}$ is in S_{Cg} . Let $W_s(v_j)$ denote the set of weak sequential limits of $\{v_j\}_{j \geq 0}$. Then, the boundedness of $\{v_j\}_{j \geq 0}$ guarantees that $W_s(v_j)$ is nonempty (see Lemma 7). Let $\hat{v} \in W_s(v_j)$; given Definition 4, it suffices to show that $\langle v - \hat{v}, A\hat{v} \rangle \geq 0 \forall v \in C$. By definition of $W_s(v_j)$, there exists a subsequence $\{v_{j_k}\}_{k \geq 0} \subseteq \{v_j\}_{j \geq 0}$ such that $v_{j_k} \rightharpoonup \hat{v}$. $\|v_j - w_j\| \rightarrow 0 \implies w_{j_k} \rightarrow \hat{v}$.

Define $G(v) = \begin{cases} Av + T_C(v), & \text{if } v \in C \\ \phi, & \text{if } v \notin C \end{cases}$, where T_C is the normality operator. According

to Lemma 9, G is maximal monotone and $0 \in G(v)$ if and only if $(v \in S_{Cg})$. Thus, it suffices to show that $(\hat{v}, 0) \in \text{graph}(G) = \{(v, u) \in \mathbb{H} \times \mathbb{H} : v \in \text{dom}(G), u \in G(v)\}$.

We recall that, for any maximal monotone operator Γ , if $\langle \Gamma x - z, x - y \rangle \geq 0 \forall x \in \text{dom } \Gamma$ and $z \in \mathbb{H}$, then $z \in \Gamma(y)$.

Thus, let $(v, u^*) \in \text{graph}(G)$ be arbitrary, $\langle u^*, v - \hat{v} \rangle \geq 0 \implies 0 \in G(\hat{v})$. Hence, we only need to verify that $\langle u^*, v - \hat{v} \rangle \geq 0$.

Now, $(v, u^*) \in \text{graph}(G) \implies u^* - Av \in T_C(v) \implies \langle u^* - Av, v - u \rangle \geq 0 \forall u \in C$.

However, by definition, $v_{j+1} = \wp_C(w_j - \lambda Av_j)$, thus, $\langle v_{j+1} - v, w_j - \lambda Av_j - v_{j+1} \rangle \geq 0 \implies \langle v - v_{j+1}, \frac{v_{j+1} - w_j}{\lambda} + Av_j \rangle \geq 0$.

Since $v_j \in C \forall j \geq 0$, we have that

$$\begin{aligned} \langle v - v_{j_k}, u^* \rangle &\geq \langle v - v_{j_k}, Av \rangle \geq \langle v - v_{j_k}, Av \rangle - \left\langle v - v_{j_k}, \frac{v_{j_k} - w_{j_k-1}}{\lambda} + Av_{j_k-1} \right\rangle \\ &= \langle v - v_{j_k}, Av - Av_{j_k} \rangle + \langle v - v_{j_k}, Av_{j_k} - Av_{j_k-1} \rangle + \frac{1}{\lambda} \langle v - v_{j_k}, w_{j_k-1} - v_{j_k} \rangle. \end{aligned}$$

Taking limit as $k \rightarrow \infty$, we get $\langle u^*, v - \hat{v} \rangle \geq 0$. Therefore, $\hat{v} \in S_{Cg}$ and $W_s(v_j) \subseteq S_{Cg}$.

Finally, we show that $v_j \rightharpoonup v' \in S_{Cg}$. To verify this, it is enough to show that $W_s(v_j)$ is a singleton. We proceed by contradiction. Suppose there exist $v', v'' \in W_s(v_j)$ with $v' \neq v''$, then there are subsequences $\{v_{j_k}\}_{k \geq 0}, \{v_{j_i}\}_{i \geq 0}$ of $\{v_j\}_{j \geq 0}$ such that $v_{j_k} \rightharpoonup v'$ and $v_{j_i} \rightharpoonup v''$. Therefore, according to Lemma 10, we have that

$$\begin{aligned} \lim_{j \rightarrow \infty} \|v_j - v'\| &= \liminf_{k \rightarrow \infty} \|v_{j_k} - v'\| < \liminf_{k \rightarrow \infty} \|v_{j_k} - v''\| = \lim_{j \rightarrow \infty} \|v_j - v''\| \\ &= \liminf_{i \rightarrow \infty} \|v_{j_i} - v''\| < \liminf_{i \rightarrow \infty} \|v_{j_i} - v'\| = \lim_{j \rightarrow \infty} \|v_j - v'\|. \end{aligned}$$

This is a contradiction since $\lim_{j \rightarrow \infty} \|v_j - v'\|$ is never less than itself.

Hence, the sequence $\{v_j\}_{j \geq 0}$ converges weakly to a critical point of g . Furthermore, if g is convex, then, from $0 \leq g(v) - g(\hat{v}) - \langle v - \hat{v}, A\hat{v} \rangle$ and $\langle v - \hat{v}, A\hat{v} \rangle \geq 0 \forall v \in C$, we obtain $g(\hat{v}) \leq g(v) \forall v \in C$. Thus, $\hat{v}$ is a minimizer of g . $\square$

Theorem 14. Let the assumptions of Theorem 12 hold for $\mathbb{H} = \mathbb{R}^n$. In addition, assume further that C is bounded and g possesses the KL property. Then, the sequence $\{v_j\}_{j \geq 0}$ generated by AMLA converges to an element of S_{Cg} .

Proof. By definition, $v_j \in C \forall j \geq 0$. Hence, the sequence $\{v_j\}_{j \geq 0}$ is bounded, and $W_s(v_j) \neq \emptyset$. We now show that $W_s(v_j) \subseteq S_{Cg}$. Let $\hat{v} \in W_s(v_j)$ be arbitrary; then, there is a subsequence $\{v_{j_k}\}_{k \geq 0}$ of $\{v_j\}_{j \geq 0}$ such that $v_{j_k} \rightarrow \hat{v}$ as $k \rightarrow \infty$. Using Theorem 12(b) and the definition of w_j , we get $\lim_{j \rightarrow \infty} \|v_{j+1} - v_j\| = \lim_{j \rightarrow \infty} \|v_{j+1} - w_j\| = \lim_{j \rightarrow \infty} \|v_j - w_j\| = 0$.

$$\text{Therefore, } \lim_{k \rightarrow \infty} v_{j_k} - w_{j_k} = \lim_{k \rightarrow \infty} \|v_{j_{k+1}} - w_{j_k}\| = \lim_{k \rightarrow \infty} \|v_{j_{k+1}} - v_{j_k}\| = 0.$$

This implies that $w_{j_k} \rightarrow \hat{v}$ and $v_{j_{k+1}} \rightarrow \hat{v}$ as $k \rightarrow \infty$. Furthermore, the continuity of A guarantees that $Av_{j_k} \rightarrow A\hat{v}$ as $k \rightarrow \infty$.

Now, let $v \in C$ be arbitrary; then, by definition of $v_{j_{k+1}}$, $\langle v_{j_{k+1}} - v, w_{j_k} - \lambda Av_{j_k} - v_{j_{k+1}} \rangle \geq 0$

$$\implies \langle v - v_{j_{k+1}}, \lambda Av_{j_k} \rangle + \langle v - v_{j_{k+1}}, v_{j_{k+1}} - w_{j_k} \rangle \geq 0.$$

Taking the limit as $k \rightarrow \infty$, we have that $\langle v - \hat{v}, \lambda A\hat{v} \rangle \geq 0 \implies \langle v - \hat{v}, A\hat{v} \rangle \geq 0 \forall v \in C$. Hence, $W_s(v_j) \subseteq S_{Cg}$.

Next, we make use of Lemma 11 to prove that $v_j \rightarrow \hat{v}$. For this, we define the map $h : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ as follows:

$$h(x, y) = g(x) + \omega \|x - y\|^2 + I_C(x) := F(x, y) + I_C(x),$$

where $I_C(x) = \begin{cases} 0, & \text{if } x \in C \\ +\infty, & \text{otherwise} \end{cases}$; then, h has the KL property on $\text{dom}(h)$.

Set $(x_j)_{j \geq 1} = (v_j, v_{j-1})_{j \geq 1}$; then, Theorem 12(a) gives that H1 of Lemma 11 holds. Moreover, the boundedness of $\{v_j\}_{j \geq 0}$ and continuity of h on C imply that there exists a subsequence $(x_{j_k})_{k \geq 1}$ of $(x_j)_{j \geq 1}$ such that $x_{j_k} \rightarrow \hat{x} = (\hat{v}, \hat{v})$ and $h(x_{j_k}) \rightarrow h(\hat{x})$, as $k \rightarrow \infty$, thus verifying H3 of Lemma 11.

Next, we show that H2 of Lemma 11 also holds. First, we recall that

$$\begin{aligned} (u_m^1, u_m^2) &= u_m \in \partial h(x_m) = (\nabla g(v_m) + 2\omega(v_m - v_{m-1}) + T_C v_m, 2\omega(v_{m-1} - v_m)) \\ &\iff u_m^1 \in \nabla g(v_m) + 2\omega(v_m - v_{m-1}) + T_C v_m \text{ and } u_m^2 = 2\omega(v_{m-1} - v_m) \\ &\iff u_m^1 - \nabla g(v_m) - 2\omega(v_m - v_{m-1}) \in T_C v_m \text{ and } u_m^2 = 2\omega(v_{m-1} - v_m) \\ &\iff u_m^1 - Av_m - 2\omega(v_m - v_{m-1}) \in T_C v_m \text{ and } u_m^2 = 2\omega(v_{m-1} - v_m) \\ &\iff \langle v_m - y, u_m^1 - Av_m - 2\omega(v_m - v_{m-1}) \rangle \geq 0 \forall y \in C \text{ and } u_m^2 = 2\omega(v_{m-1} - v_m). \end{aligned}$$

Now, by definition of v_{m+1} , we have $\langle v_{m+1} - y, w_m - \lambda Av_m - v_{m+1} \rangle \geq 0 \forall y \in C$. Thus,

$$\begin{aligned} &\left\langle v_{m+1} - y, \frac{w_m - v_{m+1}}{\lambda} - Av_m \right\rangle \geq 0 \forall y \in C \\ &\iff \left\langle v_{m+1} - y, \frac{w_m - v_{m+1}}{\lambda} + Av_{m+1} - Av_m - Av_{m+1} \right\rangle \geq 0 \forall y \in C \\ &\iff \left\langle v_{m+1} - y, \frac{w_m - v_{m+1}}{\lambda} + Av_{m+1} - Av_m + 2\omega(v_{m+1} - v_m) - Av_{m+1} - 2\omega(v_{m+1} - v_m) \right\rangle \geq 0 \forall y \in C \\ &\implies u_m = \left(\frac{w_m - v_{m+1}}{\lambda} + Av_{m+1} - Av_m + 2\omega(v_{m+1} - v_m), 2\omega(v_m - v_{m+1}) \right) \in \partial h(x_{m+1}). \end{aligned}$$

An estimation using the definition of w_m and Lipschitz continuity of A gives

$$\begin{aligned} \|u_m\| &\leq \left\| \frac{w_m - v_{m+1}}{\lambda} + Av_{m+1} - Av_m + 2\omega(v_{m+1} - v_m) \right\| + \|2\omega(v_m - v_{m+1})\| \\ &\leq \left(4\omega + K + \frac{1}{\lambda} \right) (\|v_{m+1} - v_m\| + \|v_m - v_{m-1}\|). \end{aligned}$$

Therefore, Lemma 11 guarantees that $x_j \rightarrow \hat{x} = (\hat{v}, \hat{v})$. Since $x_j = (v_j, v_{j-1})$, we have that $v_j \rightarrow \hat{v} \in W_s(v_j) \subseteq S_{Cg}$. $\square$

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Torre, L.A.; Bray, F.; Siegel, R.L.; Ferlay, J.; Lortet-Tieulent, J.; Jemal, A. Global cancer statistics, 2012. *CA Cancer J. Clin.* **2015**, *65*, 87–108. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Baskar, R.; Lee, K.A.; Yeo, R.; Yeoh, K.W. Cancer and Radiation Therapy: Current Advances and Future Directions. *Int. J. Med. Sci.* **2012**, *9*, 193–199. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Jaffray, D.A.; Gospodarowicz, M.K. Radiation Therapy for Cancer. In *Cancer: Disease Control Priorities*, 3rd ed.; Gelband, H., Jha, P., Sankaranarayanan, R., Horton, S., Eds.; The International Bank for Reconstruction and Development/The World Bank: Washington, DC, USA, 2015; Volume 3, pp. 239–247. [\[CrossRef\]](#)
5. Badey, A.; Barateau, A.; Delaby, N.; Fau, P.; Garcia, R.; De Crevoisier, R.; Lisbona, A. Overview of adaptive radiotherapy in 2019: From implementation to clinical use. *Cancer Radiother.* **2019**, *23*, 581–591. [\[CrossRef\]](#)
6. Yang, W.C.; Hsu, F.M.; Yang, P.C. Precision radiotherapy for non-small cell lung cancer. *J. Biomed. Sci.* **2020**, *27*, 82. [\[CrossRef\]](#)
7. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* **2001**, *69*, 89–95. [\[CrossRef\]](#)
8. Byrne, N.M.; Tambe, P.; Coulter, J.A. Radiation Response in the Tumour Microenvironment: Predictive Biomarkers and Future Perspectives. *J. Pers. Med.* **2021**, *11*, 53. [\[CrossRef\]](#)
9. Rückert, M.; Flohr, A.-S.; Hecht, M.; Gaip, U.S. Radiotherapy and the immune system: More than just immune suppression. *Stem Cells* **2021**, *39*, 1155–1165. [\[CrossRef\]](#)
10. Vaes, R.D.W.; Hendriks, L.E.L.; Vooijs, M.; De Ruyscher, D. Biomarkers of Radiotherapy-Induced Immunogenic Cell Death. *Cells* **2021**, *10*, 930. [\[CrossRef\]](#)
11. Brandmaier, A.; Formenti, S.C. The Impact of Radiation Therapy on Innate and Adaptive Tumor Immunity. *Semin. Radiat. Oncol.* **2019**, *30*, 139–144. [\[CrossRef\]](#)
12. Bekker, R.A.; Kim, S.; Pilon-Thomas, S.; Enderling, H. Mathematical modelling of radiotherapy and its impact on tumor interactions with the immune system. *Neoplasia* **2022**, *28*, 100796. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Keam, S.; Gill, S.; Ebert, M.A.; Nowak, A.K.; Cook, A.M. Enhancing the efficacy of immunotherapy using radiotherapy. *Clin. Transl. Immunol.* **2020**, *9*, e1169. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Finotello, F.; Trajanoski, Z. Quantifying tumor-infiltrating immune cells from transcriptomics data. *Cancer Immunol. Immunother.* **2018**, *67*, 1031–1040. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Lambrechts, D.; Wauters, E.; Boeckx, B.; Aibar, S.; Nittner, D.; Burton, O.; Bassez, A.; Decaluwé, H.; Pircher, A.; Van den Eynde, K.; et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* **2018**, *24*, 1277–1289. [\[CrossRef\]](#)
16. Petitprez, F.; Sun, C.-M.; Lacroix, L.; Sautès-Fridman, C.; de Reyniès, A.; Fridman, W.H. Quantitative Analyses of the Tumor Microenvironment Composition and Orientation in the Era of Precision Medicine. *Front. Oncol.* **2018**, *8*, 390. [\[CrossRef\]](#)
17. Cancer Genome Atlas Research Network; Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **2013**, *45*, 1113–1120. [\[CrossRef\]](#)
18. Shendure, J.; Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **2008**, *26*, 1135–1145. [\[CrossRef\]](#)
19. Sturm, G.; Finotello, F.; Petitprez, F.; Zhang, J.D.; Baumbach, J.; Fridman, W.H.; List, M.; Aneichyk, T. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics* **2019**, *35*, i436–i445. [\[CrossRef\]](#)
20. Bolis, M.; Vallerga, A.; Fratelli, M. Computational deconvolution of transcriptomic data for the study of tumor-infiltrating immune cells. *Int. J. Biol. Markers* **2020**, *35*, 20–22. [\[CrossRef\]](#)
21. Newman, A.M.; Liu, C.L.; Green, M.R.; Gentles, A.J.; Feng, W.; Xu, Y.; Hoang, C.D.; Diehn, M.; Alizadeh, A.A. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **2015**, *12*, 453–457. [\[CrossRef\]](#)
22. Newman, A.M.; Steen, C.B.; Liu, C.L.; Gentles, A.J.; Chaudhuri, A.A.; Scherer, F.; Khodadoust, M.S.; Esfahani, M.S.; Luca, B.A.; Steiner, D.; et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **2019**, *37*, 773–782. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Racle, J.; de Jonge, K.; Baumgaertner, P.; Speiser, D.E.; Gfeller, D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *eLife* **2017**, *6*, e26476. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Ai, D.; Liu, G.; Li, X.; Wang, Y.; Guo, M. Calculation of immune cell proportion from batch tumor gene expression profile based on support vector regression. *J. Bioinform. Comput. Biol.* **2020**, *18*, 2050030. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Finotello, F.; Mayer, C.; Plattner, C.; Laschober, G.; Rieder, D.; Hackl, H.; Krogsdam, A.; Loncova, Z.; Posch, W.; Wilflingseder, D.; et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* **2019**, *11*, 34. [\[CrossRef\]](#)
26. Li, B.; Severson, E.; Pignon, J.C.; Zhao, H.; Li, T.; Novak, J.; Jiang, P.; Shen, H.; Aster, J.C.; Rodig, S.; et al. Comprehensive analyses of tumor immunity: Implications for cancer immunotherapy. *Genome Biol.* **2016**, *17*, 174. [\[CrossRef\]](#)

27. Sturm, G.; Finotello, F.; List, M. Immunedecconv: An R Package for Unified Access to Computational Methods for Estimating Immune Cell Fractions from Bulk RNA-Sequencing Data. In *Bioinformatics for Cancer Immunotherapy. Methods in Molecular Biology*; Boegel, S., Ed.; Humana Press: New York, NY, USA, 2020; Volume 2120, pp. 223–232. [\[CrossRef\]](#)
28. Li, T.; Fu, J.; Zeng, Z.; Cohen, D.; Li, J.; Chen, Q.; Li, B.; Liu, X.S. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **2020**, *48*, W509–W514. [\[CrossRef\]](#)
29. Aronow, R.A.; Akbarinejad, S.; Le, T.; Su, S.; Shahriyari, L. TumorDecon: A digital cytometry software. *SoftwareX* **2022**, *18*, 101072. [\[CrossRef\]](#)
30. Avila Cobos, F.; Vandesompele, J.; Mestdag, P.; De Preter, K. Computational deconvolution of transcriptomics data from mixed cell populations. *Bioinformatics* **2018**, *34*, 1969–1979. [\[CrossRef\]](#)
31. Hao, Y.; Yan, M.; Heath, B.R.; Lei, Y.L.; Xie, Y. Fast and robust deconvolution of tumor infiltrating lymphocyte from expression profiles using least trimmed squares. *PLoS Comput. Biol.* **2019**, *15*, e1006976. [\[CrossRef\]](#)
32. Hunt, G.J.; Freytag, S.; Bahlo, M.; Gagnon-Bartsch, J.A. Dangle: Accurate and robust cell type deconvolution. *Bioinformatics* **2019**, *35*, 2093–2099. [\[CrossRef\]](#)
33. Chiu, Y.J.; Hsieh, Y.H.; Huang, Y.H. Improved cell composition deconvolution method of bulk gene expression profiles to quantify subsets of immune cells. *BMC Med. Genom.* **2019**, *12*, 169. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Dong, M.; Thennavan, A.; Urrutia, E.; Li, Y.; Perou, C.M.; Zou, F.; Jiang, Y. SCDC: Bulk gene expression deconvolution by multiple single-cell RNA sequencing references. *Brief. Bioinform.* **2021**, *22*, 416–427. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Erdmann-Pham, D.D.; Fischer, J.; Hong, J.; Song, Y.S. Likelihood-based deconvolution of bulk gene expression data using single-cell references. *Genome Res.* **2021**, *31*, 1794–1806. [\[CrossRef\]](#)
36. Hunt, G.J.; Gagnon-Bartsch, J.A. The role of scale in the estimation of cell-type proportion. *Ann. Appl. Stat.* **2021**, *15*, 270–286. [\[CrossRef\]](#)
37. Wu, Z.; Su, K.; Wu, H. Non-linear Normalization for Non-UMI Single Cell RNA-Seq. *Front. Genet.* **2021**, *12*, 612670. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Montesinos-López, A.; Montesinos-López, O.A.; Montesinos-López, J.C.; Flores-Cortes, C.A.; de la Rosa, R.; Crossa, J. A guide for kernel generalized regression methods for genomic-enabled prediction. *Heredity* **2021**, *126*, 577–596. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Mohammadi, S.; Zuckerman, N.; Goldsmith, A.; Grama, A. A critical survey of deconvolution methods for separating cell types in complex tissues. *Proc. IEEE* **2017**, *105*, 340–366. [\[CrossRef\]](#)
40. Vapnik, V.N. *The Nature of Statistical Learning Theory*, 2nd ed.; Statistics for Engineering and Information Science; Springer: New York, NY, USA, 2000; pp. 181–183. [\[CrossRef\]](#)
41. Chu, W.; Keerthi, S.S.; Ong, C.J. Bayesian support vector regression using a unified loss function. *IEEE Trans. Neural Netw.* **2004**, *15*, 29–44. [\[CrossRef\]](#)
42. Xu, Y.; Zhu, S.; Yang, S.; Zhang, C.; Jin, R.; Yang, T. Learning with Non-Convex Truncated Losses by SGD. In Proceedings of the 35th Uncertainty in Artificial Intelligence Conference, Proceedings of Machine Learning Research, Tel Aviv, Israel, 22–25 July 2019; Available online: <https://proceedings.mlr.press/v115/xu20b.html> (accessed on 15 August 2022).
43. Chen, B.; Xing, L.; Xu, B.; Zhao, H.; Zheng, N.; Principe, J.C. Kernel Risk-Sensitive Loss: Definition, Properties and Application to Robust Adaptive Filtering. *IEEE Trans. Signal Process.* **2017**, *65*, 2888–2901. [\[CrossRef\]](#)
44. Shi, W.; Xiong, K.; Wang, S. Multikernel Adaptive Filters Under the Minimum Cauchy Kernel Loss Criterion. *IEEE Access* **2019**, *7*, 120548–120558. [\[CrossRef\]](#)
45. Xu, J.; Zikatanov, L. The method of alternating projections and the method of subspace corrections in Hilbert space. *J. Am. Math. Soc.* **2002**, *15*, 573–597. [\[CrossRef\]](#)
46. Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30*, 207–210. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Le, T.; Aronow, R.A.; Kirshtein, A.; Shahriyari, L. A review of digital cytometry methods: Estimating the relative abundance of cell types in a bulk of cells. *Brief. Bioinform.* **2021**, *22*, bbaa219. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Petković, M.S.; Neta, B.; Petković, L.D.; Džunić, J. Basic concepts. In *Multipoint Methods for Solving Nonlinear Equations*; Petković, M.S., Neta, B., Petković, L.D., Džunić, J., Eds.; Academic Press: Oxford, UK, 2013; pp. 1–26. [\[CrossRef\]](#)
49. Arora, J.S. Numerical Methods for Constrained Optimum Design. In *Introduction to Optimum Design*, 3rd ed.; Arora, J.S., Ed.; Academic Press: Oxford, UK, 2012; pp. 491–531. [\[CrossRef\]](#)
50. Atallah, M.J.; Chen, D.Z. Deterministic Parallel Computational Geometry. In *Handbook of Computational Geometry*; Sack, J.R., Urrutia, J., Eds.; North-Holland: Amsterdam, The Netherlands, 2000; pp. 155–200. [\[CrossRef\]](#)
51. Attouch, H.; Bolte, J.; Redont, P.; Soubeyran, A. Proximal Alternating Minimization and Projection Methods for Nonconvex Problems: An Approach Based on the Kurdyka-Łojasiewicz Inequality. *Math. Oper. Res.* **2010**, *35*, 438–457. Available online: <http://www.jstor.org/stable/40801236> (accessed on 17 August 2022). [\[CrossRef\]](#)
52. László, S.C. Convergence rates for an inertial algorithm of gradient type associated to a smooth non-convex minimization. *Math. Program.* **2021**, *190*, 285–329. [\[CrossRef\]](#)
53. Khamaru, K.; Wainwright, M.J. Convergence guarantees for a class of non-convex and non-smooth optimization problems. *J. Mach. Learn. Res.* **2019**, *20*, 1–52. Available online: <https://jmlr.org/papers/volume20/18-762/18-762.pdf> (accessed on 19 August 2022).

-
54. Eberlein, W.F. Weak Compactness in Banach Spaces I. *Proc. Natl. Acad. Sci. USA* **1947**, *33*, 51–53. Available online: <http://www.jstor.org/stable/87813> (accessed on 5 August 2022). [[CrossRef](#)]
 55. Chidume, C.E. *Geometric Properties of Banach Spaces and Nonlinear Iterations*; Springer: London, UK, 2009; pp. 75–76. [[CrossRef](#)]
 56. Rockafellar, R.T. On the Maximality of Sums of Nonlinear Monotone Operators. *Trans. Am. Math. Soc.* **1970**, *149*, 75–88. [[CrossRef](#)]
 57. Opial, Z. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Am. Math. Soc.* **1967**, *73*, 591–597. [[CrossRef](#)]
 58. Ochs, P.; Chen, Y.; Brox, T.; Pock, T. iPiano: Inertial proximal algorithm for nonconvex optimization. *SIAM J. Imaging Sci.* **2014**, *7*, 1388–1419. [[CrossRef](#)]