Bias in allocations using differentially private census data: An analysis of the 2020 U.S. Census

Author One[a,c,1], **Author Two**[b,1,2], **and Author Three**[a]

[a]Affiliation One; [b]Affiliation Two; [c]Affiliation Three

This manuscript was compiled on February 23, 2023

The decennial census is a primary data for the US government to make critical decision. For example, 132 programs used Census Bureau data to distribute more than $675 billion in funds during fiscal year 2015. In order to ensure the published census data does not reveal individual information, Census Bureau adopts the differential privacy (DP) technique. In particular, in 20202 the Census Bureau implemented the Top Down algorithm which release statistical data from top (nation) to bottom hierarchical level (block). However, it was recently observed that the DP outcomes can introduce biased outcomes, especially for the minority groups. In this paper, we analyzes the reasons for these disproportionate impacts and proposes guidelines to mitigate these effects. We focus on two aspects that can produce the fair outcomes: (1) shape of allocation function and (2) impact of post-processing steps.

Keyword 1 | Keyword 2 | Keyword 3 | ...

Introduction

Agencies, such as the U.S. Census Bureau, release data sets and statistics about groups of individuals that are then used as inputs to a number critical decision processes. For example, the census data is used to decide whether a jurisdiction must provide language assistance during elections, Title I fund allocation in education (**?**) and to establish national level COVID-19 vaccination distribution plans for states and jurisdictions (**?**). The resulting decisions can have significant societal, economic, and medical impacts for participating individuals.

In many cases, the released data contain sensitive information and their privacy is strictly regulated. For example, in the U.S., the census data is regulated under Title 13 (**?**), which requires that no individual be identified from any data release by the Census Bureau. In Europe, data release are regulated according to the *General Data Protection Regulation* (**?**), which addresses the control and transfer of personal data.

Statistical agencies thus release *privacy-preserving* data and statistics that conform to privacy and confidentiality requirements. In the U.S., a small number of decisions, such as congressional apportionment, are taken using unprotected true values, but the vast majority of decisions rely on privacy-preserving data. Of particular interest are resource allocation decisions relying on the U.S. Census Bureau data, since the bureau will release several privacy-preserving data products using the framework of *Differential Privacy* (**?**) for their 2020 release. In particular, in 2020 the Census Bureau implemented a new privacy preserving framework to release privately hierarchical statistical data, the Top Down algorithm. The algorithms works by firstly splitting the given privacy budget ϵ to six hierarchical levels (nation, state, county, tract, block, group—block). Then in the second step, a post-processing step is applied to make sure the noisy counts are consistent, e.g., the counts should be non-negative. However, (**?**) empirically showed that differential privacy may have a disparate impact on several resource allocation problems. The noise introduced by the privacy mechanism may result in decisions that impact various groups differently. Unfortunately, the paper did not provide a deep understanding why this behavior happens and any mitigation to resolve the issue. This paper builds on these observations and provides a step towards a deeper understanding of the fairness issues arising when differentially private data is used as input to several resource allocation problems. *One of its main results is to prove that several allotment problems and decision rules with significant societal impact (e.g., the allocation of educational funds, the decision to provide minority language assistance on election ballots, or the distribution of COVID-19 vaccines) exhibit inherent unfairness when applied to a differentially private release of the census data.* To counteract this negative results, the paper examines the conditions under which decision making is fair when using differential privacy, and techniques to bound unfairness. The paper also provides a number of mitigation approaches to alleviate biases introduced by differential privacy on such decision making problems. More specifically, the paper makes the following contributions:

1. It formally defines notions of fairness and bounded fairness for decision making subject to privacy requirements.
2. It examines the roots of the induced unfairness by analyzing the structure of the decision making problems.
3. It proposes several guidelines to mitigate the negative fairness effects of the decision problems studied.

To the best of the authors' knowledge, this is the first study that attempt at characterizing the relation between differential privacy and fairness in decision problems. All proofs are reported in the appendix.

Significance Statement

Authors must submit a 120-word maximum statement about the significance of their research paper written at a level understandable to an undergraduate educated scientist outside their field of speciality. The primary goal of the significance statement is to explain the relevance of the work in broad context to a broad readership. The significance statement appears in the paper itself and is required for all research papers.

Please provide details of author contributions here.

Please declare any competing interests here.

[1]A.O.(Author One) contributed equally to this work with A.T. (Author Two) (remove if not applicable).

[2]To whom correspondence should be addressed. E-mail: author.twoemail.com

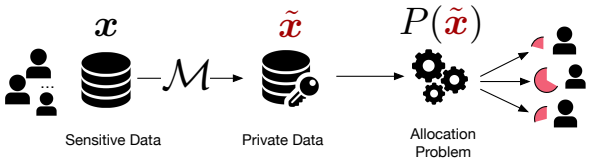

Fig. 1. Diagram of the private allocation problem.

Preliminaries: Differential Privacy

Differential Privacy (**?**) (DP) is a rigorous privacy notion that characterizes the amount of information of an individual's data being disclosed in a computation.

Definition 1 *A randomized algorithm $\mathcal{M} : \mathcal{X} \to \mathcal{R}$ with domain $\mathcal{X}$ and range $\mathcal{R}$ satisfies* ϵ-differential privacy *if for any output $O \subseteq \mathcal{R}$ and data sets $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ differing by at most one entry (written $\boldsymbol{x} \sim \boldsymbol{x}'$)*

$$\Pr[\mathcal{M}(\boldsymbol{x}) \in O] \leq \exp(\epsilon) \Pr[\mathcal{M}(\boldsymbol{x}') \in O]. \quad [1]$$

Parameter $\epsilon > 0$ is the *privacy loss*, with values close to 0 denoting strong privacy. Intuitively, DP states that any event occur with similar probability regardless of the participation of any individual data to the data set.

DP satisfies several properties including *immunity to post-processing*, which states that the privacy loss of DP outputs is not affected by arbitrary data-independent post-processing (**?**).

A function f from a data set $\boldsymbol{x} \in \mathcal{X}$ to a result set $R \subseteq \mathbb{R}^n$ can be made differentially private by injecting random noise onto its output. The amount of noise relies on the notion of *global sensitivity* $\Delta_f = \max_{\boldsymbol{x} \sim \boldsymbol{x}'} \|f(\boldsymbol{x}) - f(\boldsymbol{x}')\|_1$. The *Laplace mechanism* (**?**) that outputs $f(\boldsymbol{x}) + \boldsymbol{\eta}$, where $\boldsymbol{\eta} \in \mathbb{R}^n$ is drawn from the i.i.d. Laplace distribution with 0 mean and scale Δ_f/ϵ over n dimensions, achieves ϵ-DP.

Problem Setting and Goals

The paper considers a dataset $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^k$ of n entities, whose elements $x_i = (x_{i1}, \ldots, x_{1k})$ describe k measurable quantities of entity $i \in [n]$, such as the number of individuals living in a geographical region i and their English proficiency. The paper considers two classes of problems:

- An *allotment problem* $P : \mathcal{X} \times [n] \to \mathbb{R}$ is a function that distributes a finite set of resources to some problem entity. P may represent, for instance, the amount of money allotted to a school district.
- A *decision rule* $P : \mathcal{X} \times [n] \to \{0, 1\}$ determines whether some entity qualifies for some benefits. For instance, P may represent if election ballots should be described in a minority language for an electoral district.

The paper assumes that P has bounded range, and uses the shorthand $P_i(\boldsymbol{x})$ to denote $P(\boldsymbol{x}, i)$ for entity i. The focus of the paper is to study the effects of a DP data-release mechanism $\mathcal{M}$ to the outcomes of problem P. Mechanism $\mathcal{M}$ is applied to the dataset $\boldsymbol{x}$ to produce a privacy-preserving counterpart $\tilde{\boldsymbol{x}}$ and the resulting private outcome $P_i(\tilde{\boldsymbol{x}})$ is used to make some allocation decisions.

Figure 1 provides an illustrative diagram.

Because random noise is added to the original dataset $\boldsymbol{x}$, the output $P_i(\tilde{\boldsymbol{x}})$ incurs some error. *The focus of this paper is to characterize and quantify the disparate impact of this error among the problem entities.* In particular, the paper focuses on two notations of errors.

Definition 2 (Statistical bias) *The statistical bias $B_P^i(\mathcal{M}, \boldsymbol{x})$ of the mechanism $\mathcal{M}$ measures the difference between the expected private outcome with the true outcome:*

$$B_P^i(\mathcal{M}, \boldsymbol{x}) = \mathbb{E}_{\tilde{\boldsymbol{x}} \sim \mathcal{M}(\boldsymbol{x})} [P_i(\tilde{\boldsymbol{x}})] - P_i(\boldsymbol{x}), \quad [2]$$

The paper also considers another notation of error which is the normalized version of the above bias.

Definition 3 (Multiplicative error) *The multiplicative error under mechanism $\mathcal{M}$ and problem P for entity i is given by: $B_P^i(\mathcal{M}, \boldsymbol{x}) / P^i(x)$*

Our notation of fairness will be based on these two notations of errors.

Definition 4 (α-fairness (?**))** *Given the true data $\boldsymbol{x}$, the mechanism $\mathcal{M}$ is said to be* α-fair *if, for any $i \in [n]$,*

$$\xi_P^i(\mathcal{M}, \boldsymbol{x}) = \left| B_P^i(\mathcal{M}, \boldsymbol{x}) - B_P^j(\mathcal{M}, \boldsymbol{x}) \right| \leq \alpha,$$

where $\xi_P^i(\mathcal{M}, \boldsymbol{x})$ is referred to as the disparity error *associated with district i. The mechanism $\mathcal{M}$ is α'-*minimally fair *if $\alpha' = \inf \alpha$ such that $\mathcal{M}$ is α-fair. To put it differently, the mechanism $\mathcal{M}$ is α'-*minimally fair *if*

$$\begin{aligned} \alpha' &= \max_{j \neq i} \left| B_P^i(\mathcal{M}, \boldsymbol{x}) - B_P^j(\mathcal{M}, \boldsymbol{x}) \right| \\ &= \max_{j \in [n]} B_P^j(\mathcal{M}, \boldsymbol{x}) - \min_{j \in [n]} B_P^j(\mathcal{M}, \boldsymbol{x}). \end{aligned}$$

Throughout this report, every time we say that a mechanism is α-fair, we mean that it is α-minimally fair.

Motivation examples

A. Title I school allocation. The *Title I of the Elementary and Secondary Education Act of 1965* (**?**) distributes about $6.5 billion through basic grants. The federal allotment is divided among nearly 17000 qualifying school districts in proportion to the count x_i of children aged 5 to 17 who live in necessitous families in district i. The allocation is formalized by:

$$P_i^F(\boldsymbol{x}) \stackrel{\text{def}}{=} \left(\frac{x_i \cdot a_i}{\sum_{i \in [n]} x_i \cdot a_i} \right), \quad [3]$$

where $\boldsymbol{x} = (x_i)_{i \in [n]}$ is the vector of all districts counts and a_i is a weight factor reflecting students expenditures. We consider applying Laplace mechanism with two privacy budgets $\epsilon = 0.01$ and $\epsilon = 0.001$ to release privately the school' population. These private counts are used to determine amount of money each school district should receive. The statistical bias in Definition 2 here represents the gain/loss in money USD a school can have under the privacy preserving mechanism. Figure 2 summarizes such quantity for all school districts in New York state.

It can be seen from Figure 2 that schools of low population receive much more money than their actual needs, while schools of high population receive less. This figure made a clear evidence on the disparate impact of the privacy preserving mechanism in practice.

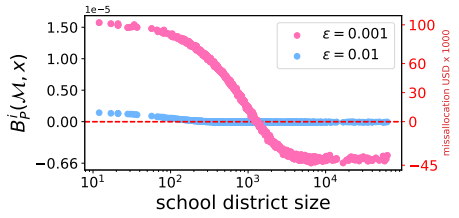

Fig. 2. Disproportionate Title 1 Funds Allocation in NY.

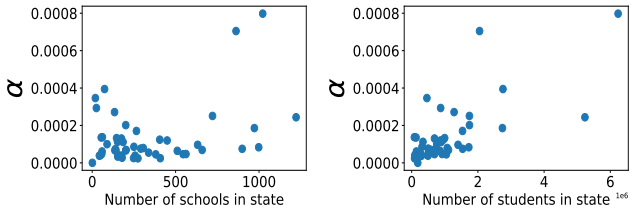

Fig. 3. Correlation between number of schools (left) and number of students (right) per state with its level of α fairness.

Further analysis. While there is a some level of unfairness under privacy-preserving mechanism towards different school districts, we observed that level varies differently by state. We report the fairness level per state, i.e the difference between the maximum bias minus the minimum bias per state in Figure 4. We see that under $\epsilon = 0.1$ California has the largest unfairness level over its schools.

To understand why California has the highest level of unfairness, we provide the scatter plot between level of fairness α per state with its number of schools and its number of students in Figure 3. We observed a Pearson correlation of 0.37 between the number of schools per state and its fairness level. Also, a high Pearson correlation of 0.71 between number of state's students and its fairness level . A In other words, the state with more schools and more students tends to be suffered more unfairness. For example, the California state has the highest number of students, more than 6.5 million students overall. It also ranks secondly based on the number of schools with 1022 schools just below Texas with 1222 schools. This partly explains why this state suffers the most unfairness for all schools as showed in Figure 4.

Differentially Private mechanisms

To release privately the outcomes P_i^F given a privacy constraint ϵ, there are several mechanisms. These mechanisms can roughly be divided into the two categories, strict and non-strict allocation mechanisms. A strict allocation mechanism requires that its outcome should always lie in the probability simplex $\Delta_n = \{x \,|x \in \mathbb{R}^{+n}, \mathbf{1}^T x = 1\}$ while a non-strict allocation mechanism only asks its output to be non-negative. The rest of this report aims to study the (approximate) optimal (strict allocation) mechanisms under different fairness metrics.

Strict Allocation Mechanism.

Definition 5 (Baseline Mechanism (BL)) *The* baseline mechanism *outputs the allocation for each distribute $i \in [n]$ as follows.*

$$\mathcal{M}_{\mathrm{BL}}\left(\tilde{\boldsymbol{x}}\right)_i = \frac{a_i \cdot \left(\tilde{x}_i\right)_+}{\sum_{j=1}^n a_j \cdot \left(\tilde{x}_j\right)_+}\,.$$

Where $\tilde{x}_i$ is the noisy private population count, while the supscript $x_+ = \max(x, 0)$ takes the non-negative part of the number x.

Definition 6 (Projection onto Simplex Mechanism (PoS)) *The* projection onto simplex mechanism *outputs the allocation for each distribute $i \in [n]$ as follows.*

$$\mathcal{M}_{\mathrm{PoS}}\left(\tilde{\boldsymbol{x}}\right)_i = \underset{\boldsymbol{v}\in\mathbb{R}^n}{\arg\min} \left\|\boldsymbol{v} - P^F\left(\tilde{\boldsymbol{x}}\right)\right\|_2 \qquad \text{s.t. } \sum_{i=1}^n v_i = 1,\ \boldsymbol{v} \geq \mathbf{0}\,.$$

Non-strict Allocation Mechanism.

Definition 7 (Positive Allocation Mechanism (PA)) *The* positive allocation mechanism *outputs the allocation for each distribute $i \in [n]$ as follows.*

$$\mathcal{M}_{\mathrm{PA}}\left(\tilde{\boldsymbol{x}}\right)_i = \left(P_i^F\left(\tilde{\boldsymbol{x}}\right)\right)_+ = \left(\frac{a_i \cdot \tilde{x}_i}{\sum_{j=1}^n a_j \cdot \tilde{x}_j}\right)_+ .$$

Definition 8 (Repair Mechanism (RP) (?)) *The* repair mechanism *outputs the allocation for each distribute $i \in [n]$ as follows.*

$$\mathcal{M}_{\mathrm{RP}}\left(\tilde{\boldsymbol{x}}\right)_i = \frac{a_i \cdot \left(\tilde{x}_i\right)_+ + \Delta}{\sum_{j=1}^n a_j \cdot \left(\tilde{x}_j\right)_+ - \Delta'}\,,$$

where

$$\Delta = \frac{\ln\left(2n/\delta\right)}{\epsilon}\,, \qquad \Delta' = \frac{n \ln\left(2n^2/\delta\right)}{\epsilon}\,.$$

Proposition 1 (No-penalty allocation (?)) *The following inequality holds with probability at least $1 - \delta$.*

$$\mathcal{M}_{\mathrm{RP}}\left(\tilde{\boldsymbol{x}}\right)_i \geq P_i^F\left(\boldsymbol{x}\right)\,, \qquad \forall\ i \in [n]\,.$$

Source of unfairness

We investigate the two main sources of unfairness highlighted in previous section: (1) shape of allocation function and (2) post-processing steps.

Shape of allocation function.

Theorem 1 *Let P be an allotment problem which is at least twice differentiable. A data-release mechanism $\mathcal{M}$ is α-fair w.r.t. P for some $\alpha < \infty$ if there exist some constant values c_{jl}^i $(i \in [n], j, l \in [k])$ such that, for all datasets $\boldsymbol{x} \in \mathcal{X}$,*

$$(\boldsymbol{H}P_i)_{j,l}(\boldsymbol{x}) = c_{j,l}^i \quad (i \in [n]\ j, l \in [k]).$$

Corollary 1 *If P is a linear function, then $\mathcal{M}$ is fair w.r.t. P.*

Corollary 2 *$\mathcal{M}$ is fair w.r.t. P if there exists a constant c such that, for all dataset $\boldsymbol{x}$,*

$$Tr(\boldsymbol{H}P_i)(\boldsymbol{x}) = c \quad (i \in [n]).$$

Corollary 3 *Consider an allocation problem P. Mechanism $\mathcal{M}$ is not fair w.r.t. P if there exist two entries $i, j \in [n]$ such that $Tr(\boldsymbol{H}P_i)(\boldsymbol{x}) \neq Tr(\boldsymbol{H}P_j)(\boldsymbol{x})$ for some dataset $\boldsymbol{x}$.*

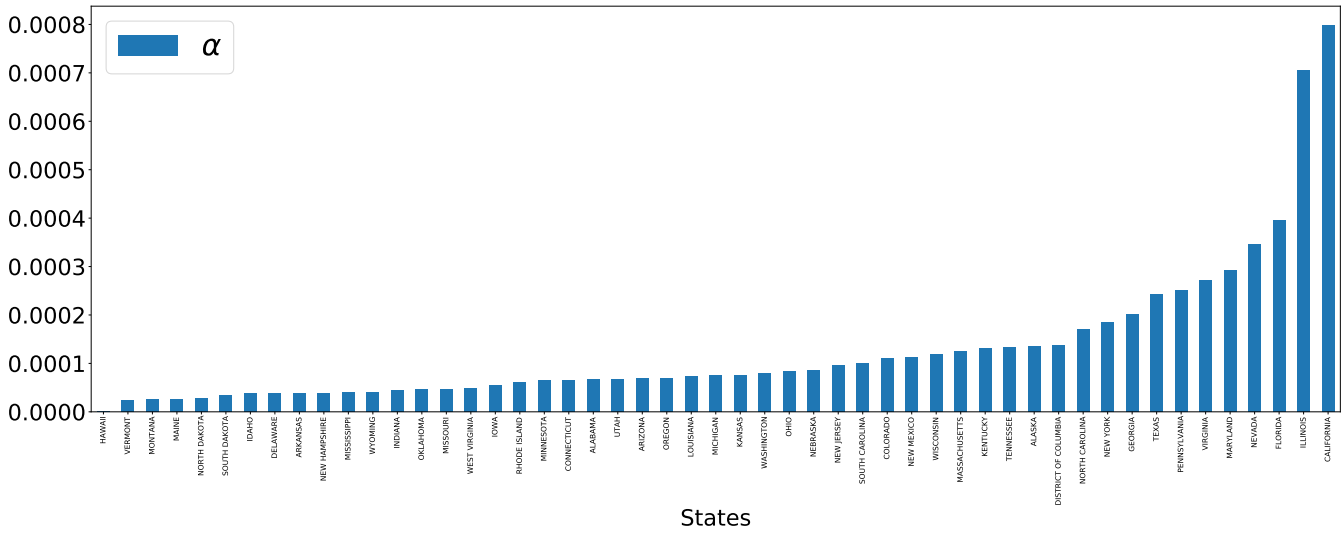

Fig. 4. Level of α fairness per state under Baseline mechanism (BL) at $\epsilon = 0.1$

The above implies that fairness cannot be achieved if P is *a non-convex function*, as is the case for *all* the allocation problems considered in this paper. *A fundamental consequence of this result is the recognition that adding Laplacian noise to the inputs of the motivating example will necessarily introduce fairness issues.* For instance, consider P^F and notice that the trace of its Hessian

$$\mathrm{Tr}(\boldsymbol{H}P_i^F) = 2a_i \left[\frac{x_i \sum_{j\in[n]} a_j^2 - a_i \left(\sum_{j\in[n]} x_j a_j \right)}{\left(\sum_{j\in[n]} x_j a_j \right)^3} \right],$$

is not constant with respect to its inputs. Thus, any two entries i, j whose $x_i \neq x_j$ imply $\mathrm{Tr}(\boldsymbol{H}P_i^F) \neq \mathrm{Tr}(\boldsymbol{H}P_j^F)$. As illustrated in Figure 2, Problem P^F can introduce significant disparity errors. For $\epsilon = 0.001, 0.01$, and 0.1 the estimated fairness bounds are 0.003, 3×10^{-5}, and 1.2×10^{-6} respectively, which amount to an average misallocation of \$43,281, \$4,328, and \$865.6 respectively. The estimated fairness bounds were obtained by performing a linear search over all n school districts and selecting the maximal $\mathrm{Tr}(\boldsymbol{H}P_i^F)$.

Impact of post-processing. The post-processing steps can be applied at the inputs x, over the outcome $P_i^F(x)$ or at both. We investigate the impact of post-processing over input and outcome separately in this section.

Post-processing over the inputs. This step is performed to make sure the released private counts satisfies consistency constraints (**?**). For example, the released private counts should be non-negative integer numbers, or sum of counts at all cities' in a state should be equal to that state's count.

Non-negative truncation $\tilde{x} = \max(0, \tilde{x})$

We have the following result which state that non-negative truncation introduces positive bias, and the closer to zero the true count is, the higher the bias.

Theorem 2 *Let* $\tilde{x} = x + Lap(\lambda)$*, with scale* $\lambda > 0$*, and* $\hat{x} = PP^{\geq \ell}(\tilde{x})$*, with* $\ell < x$*, be its post-processed value. Then,*

$$\mathbb{E}[\hat{x}] = x + \frac{\lambda}{2} \exp(\frac{\ell - x}{\lambda}).$$

Integral transform

The integral transform $\mathrm{PP}^{\mathbb{N}}(z)$ is used when the released data should be of integral quantities. To make sure that this processing step does not introduce additional bias, we can rely on the stochastic rounding technique:

$$\mathrm{PP}^{\mathbb{N}}(z) = \begin{cases} \lfloor z \rfloor \text{ w.p.: } 1 - (z - \lfloor z \rfloor) \\ \lfloor z \rfloor + 1 \text{ w.p.: } z - \lfloor z \rfloor \end{cases} \quad [4]$$

The stochastic rounding guarantees that $\mathbb{E}[\mathrm{PP}^{\mathbb{N}}(\tilde{x})] = \tilde{x}$ so no additional bias will introduce to $\mathrm{PP}^{\mathbb{N}}(\tilde{x})$

Post-processing over the outcomes.

Mitigating Solutions

Mechanisms. Different kind of post-processing mechanisms are considered in this section. These mechanisms require that their outcomes should always lie in the probability simplex Δ_n. The rest of this work aims to study the (approximate) optimal mechanisms under different fairness metrics.

Definition 9 (Baseline Mechanism (BL)) *The* baseline mechanism *outputs the allocation for each entity* $i \in [n]$ *as follows.*

$$\mathcal{M}_{\mathrm{BL}}(\tilde{\boldsymbol{x}})_i = \frac{a_i \cdot (\tilde{x}_i)_+}{\sum_{j=1}^n a_j \cdot (\tilde{x}_j)_+}.$$

Definition 10 (Projection onto Simplex Mechanism (PoS)) *The* projection onto simplex mechanism *outputs the allocation for each entity* $i \in [n]$ *as follows.*

$$\mathcal{M}_{\mathrm{PoS}}(\tilde{\boldsymbol{x}})_i = \arg\min_{\boldsymbol{v}\in\mathbb{R}^n} \left\| \boldsymbol{v} - P^F(\tilde{\boldsymbol{x}}) \right\|_2 \qquad \text{s.t. } \sum_{i=1}^n v_i = 1,\ \boldsymbol{v} \geq \boldsymbol{0}.$$