Evolving Fair Linear Regression for the Representation of Human-Drawn Regression Lines

Mario Köppen and Kaori Yoshida and Kei Ohnishi

Abstract— **Here we study a generalization of linear regression to the case of maximal elements of a general fairness relation. The regression then is based on balancing the distances to the data points. The studied relations are lexicographic minimum, maxmin fairness, proportional fairness, and majorities, all in a complementary version to represent minimality. A new combination of proportional fairness and majority is introduced as well. Experiments are performed on human subjects solving the visual task to draw a line fitting to given data points, and by use of evolutionary computation (here by Differential Evolution) the weights of a fair linear regression are adjusted to the human-provided results. The fact that this gives a more precise approximation than (weighted) linear regression hints on the inclusion of the balance among the distances to the given data points in the human decision making process.**

I. Introduction

Among the various regression tasks, linear regression seems to be a solved problem. The standard equations for linear regression are based on the minimization of the total sum of shortest distances of a regression line to the given data points. In addition to an analytic expression that can solve the optimization problem, the formula has low complexity and is provided in any standard statistical software. In fact, there might be only a few contexts where the assumptions that lead to standard linear regression are not valid. However, they exists. Here we give two thoughts that motivate the need for an expansion of the concept of linear regression. (1) A common study problem in mathematical economics is the distribution of goods among a number of (human) agents, as well as the sharing of resources. Such problems become of high relevance with the growth of the internet and related sharing problems. There are cases where global optimization leads to solutions that exclude users from participation in the network utilization (see e.g. [1] or [2]). To overcome this problem, various extensions of the concept of optimality have been introduced, usually put together under the term "fairness." We can consider that also for the cognitive task of fitting a line by a human observer, i.e. a task that involves a human decision maker, the issue of fairness can become relevant. Instead of the "utilitarian" point of view to sum up deviations from the given data points, as it is done in linear regression, the focus is given to the distribution of deviations, judged against each other: who gets more, who gets less? (2) The linear regression is based on a pure geometric point of view, by using Euclidian distances between points and lines. Also here, if a human decision maker is involved, the focus will rather be on mutual losses and wins [3]. At least mean absolute error then appears more appropriate (see also [4] for a related discussion) than Euclidian distance — but there is no analytic derivation for a corr. extreme value problem and the approach is not further considered.

Mario Köppen is with the Graduate School of Creative Informatics, Kaori Yoshida with the Brain Sciences Dept. and Kei Ohnishi with the Computer Science and Electrical Eng. Dept. all of Kyushu Institute of Technology, Fukuoka, Japan (email: mkoeppen@ci.kyutech.ac.jp, kaori@brain.kyutech.ac.jp, ohnishi@cse.kyutech.ac.jp).

Here we want to present a general approach to linear regression that includes standard linear regression as a special case. The approach is based on maximality of set-theoretic binary relations. Moreover, we report on the results of a (alas small in scope) experiment that demonstrates the feasibility of having alternative linear regression concepts at hand. The task is related to a human visual cognition task. But due to underlying complexity, instead of an analytical solution we will need an optimization method to solve the related optimization problems (Differential Evolution is used in our case). The fair linear regression will be presented in the next section. Section III than provides the report on the visual task of fitting regression lines by a human observer. Section IV gives a summary of the paper.

II. Fair Regression

The starting point is the pairwise comparison between items of a domain. It comes out that this is all what is needed to specify an optimization tasks. Formally, the domain is a set A and the pairwise comparison outcome for any pair (x, y) of elements of the domain gives a binary relation R. Thus, in so-called set representation, a relation R is a subset of $A \times A$, i.e. the set of ordered pairs of elements of A.

In general, binary relations exhibit two main aspects of comparison between elements: one aspect can be roughly described as equality, similarity, nearness etc. the other aspect as preference, betterness, or advance. In fact, any relation can be split into two disjoint relations called the symmetric part and the asymmetric part. The symmetric part $I(R)$ of a relation R contains all pairs $(x, y) \in R$ where also $(y, x) \in R$. On the contrary, the asymmetric part $P(R)$ contains all pairs $(x, y) \in R$ where (y, x) does not belong to R. The asymmetric part represents the comparison aspect of a relation.

Moreover, the asymmetric part allows for a general definition of extremity of elements of the domain A. We consider all elements x of A where there is no $y \in A$ such that $(y, x) \in P(R)$ as maximal elements of R, establishing the maximum set. In a similar way we can define minimal elements x where there is no y such that $(x, y) \in P(R)$ [5].

It can be easily seen that for finite non-empty domains A a necessary and sufficient condition for a relation to have a non-empty maximum (or minimum set) is cycle-freeness of the asymmetric part. Cycle-freeness refers to the representation of a relation as a directed graph and means that the graph of $P(R)$ does not contain any cycles. Formally it means that there is no set x_i with $i = 1, \ldots, k$ of $k \geq 3$ elements such that $(x_1, x_2) \in P(R), (x_2, x_3) \in P(R), \ldots, (x_{k-1}, x_k) \in P(R)$ and $(x_k, x_1) \in P(R)$. The condition for $k = 1, 2$ is automatically fulfilled by the definition of the asymmetric part $P(R)$ that will not contain any pair (x_1, x_1) as well as any pair (x_2, x_1) whenever $(x_1, x_2) \in R$.

A small comment here: it is also possible to consider greatest elements x (or least elements in a similar way) where $(x, y) \in R$ for all $y \in A$. However, it is more ambitious than maximality and for many specifications of relations R over domains A there are no such x at all. Thus, in the following the focus will be on maximal elements, i.e. essentially elements to which no other element of the domain is in relation, as well as relations where the main aspect is comparability and not equality. We will use an in-fix notation $x \geq_R y$ to express the fact that $(x, y) \in R$ and the notation $x >_R y$ to express $(x, y) \in P(R)$.

The task of relational optimization then is to establish a relation R that captures the optimality aspect of a model domain, and to find its maximum set as solution. This definition is general enough to cover the classical approach of functional optimization (where the real-valued larger-relation serves as relation) as well as multi-objective optimization, where the Pareto-dominance relation is used (and in which case the maximum set is called Pareto front).

One important class of relations that can be used to paraphrase optimization goals are fairness relations [2]. These class of relations represents aspects of the distribution or sharing of resources when comparing solutions. It is convenient to focus hereby on relations that are cycle-free (so they have non-empty maximum sets) and are implied by Pareto-dominance (so their maximum sets are Pareto-efficient solutions).

Here we want to apply the relational optimization approach to the linear regression problem. Given a set of N data points $(x_i, y_i), i = 1, \ldots, N$, where the x_i and y_i are from R^2 we are seeking the two parameters slope a and bias b of a line according to the equation

$$y = ax + b \tag{1}$$

that in some sense best fits to the data. For the common "textbook" linear regression this is the mean squared error (MSE) of the Euclidian distances of the points to the line (distance here means the shortest distance of the point to a point of the line). In general, it will be a vector whose components represents the deviation from the line that is not necessarily the Euclidian distance. Now we have to consider two aspects:

1) Choice of a suitable set of relations: as mentioned before, fairness relations appear to be of interest since they can represent the balance of an allocation in an (Pareto-)efficient manner. Since the base are vectors, these should be relations whose domain are subsets of R_+^N. Moreover, these relations should fit with the possible numerical range of a regression problem (especially the handling of vectors with component values 0 or values close to 0). Also, their maximum sets should preferably contain only few (or even only one) elements to simplify the selection of a final solution.
2) In resource allocation tasks, the problems are usually formulated in terms of maximality. For example (it will be discussed later in more detail) the approach of seeking a vector with largest minimal component is assuming an advantage of *increasing* components. However, in regression we have to consider minimality: the closer all points to the regression line the better. On first glance it seems no major problems to rephrase a relation with maximality aspect to a minimality one. The converse relation of a relation R (composed of all pairs (x, y) where $(y, x) \in R$) will serve maximum sets that are equal to the minimum set of R. But in many cases this approach does not give a suitable relation, basically because there is no canonical and domain-independent level of aspiration for minimality as it is "zero allocation avoidance" for maximality — that would be only a non-numerical infinite level and otherwise depending on extreme values of the domain. Therefore, it appears more of practical value to change a maximality relation in a syntactic way to specify minimality relations, despite the fact that then the new relations are not formally related to the former ones. We will speak about the *formal complement* if using a converse relation, and *syntactic complement* if we base the definition of a relation on complementing to syntactic elements of its definition: replace maximum by minimum, larger by smaller etc.

In the following, we will discuss possible (maximality) fairness relations and their potential to provide related minimality relations that can be used for the linear regression problem. Before continuing, we need to provide basic definitions.

Definition 1: Given a vector domain $A \subseteq R^n$ then the vector x_i *Pareto-dominates* the vector y_i if for all $i = 1, \ldots, n$ $x_i \geq y_i$ and for at least one j from same range as i it is $x_j > y_j$.

Definition 2: Lexicographic order: when comparing any two vectors x and y from R_n it is said that x_i comes *lexicographically before* y if there is an index j from $\{1, 2, \ldots, n\}$ such that for $1 \leq k < j$ $x_k = y_k$ and $x_j > y_j$.
It means that like sorting books in a library by title, first comparison is made by the first component of both vectors, and in case they are equal by second component etc.

Definition 3: A *sorting permutation* of an n-dimensional vector x is any permutation Π of the vector indices such that for $1 \leq i < j \leq n$ it is $x_i \leq x_j$. The i-th component of the permuted vector is usually denoted as $x_{(i)}$.
Note that a vector can have more than one sorting per-

mutation, in case there are repeated occurrences of the same numerical value. Care has to be taken that this will not influence definitions based on $x_{(i)}$-components, like the following one (where it does not matter which permutation is chosen):

Definition 4: Given an n-dimensional vector x then the i-th component of its majorant vector is given by $\sum_{k=1}^{i} x_{(k)}$, i.e. its first component is the smallest component of x, the second component is the sum of the smallest and second-smallest components of x etc. and the n-th component is the total sum of all components of x.

Definition 5: Given is a weight vector w from R^n. The ordered weighted averaging (OWA) of a vector $x \in R^n$ by w is computed as

$$owa_w(x) = \sum_{i=1}^{n} w_i x_{(i)} \tag{2}$$

Now we can discuss the suitability of a number of relations and see whether they can serve as substitute relations for the "default" minimal MSE criterion in linear regression in a "fair sense".

A. Pareto dominance

We start with Pareto dominance but just to confirm that this relation is not suitable for at least two reasons. The minimality version of Pareto-dominance is the same, no matter if we choose the formal or the syntactic approach: for $x <_P y$ we require that all $x_i \leq y_i$ and at least one $x_j < y_j$ ($1 \leq i, j \leq n$).

1) The size of maximum sets of the Pareto dominance relation is often even a larger share of the domain. With increasing dimension of the domain (as it would be the number of data points here) the chance that two vectors are in this relation to each other is known to fall exponentially, which causes a smaller number of cases where some y is in relation to a given x (it would need to be larger in *all* components) and thus any x more likely to belong to the maximum set.
2) A line passing through any pair of data points would be Pareto-optimal: the line is uniquely specified and the deviation vector would contain two 0s for which no value can be smaller.

B. Maxmin fairness

Maxmin fairness was established in the seminal work [6] in networking research and has gained increasing attention since then. Maxmin fairness actually represents the characteristic of the final result of an algorithm "Bottleneck flow control" (BFC) [7] but has found variants of specifications since then. The BFC algorithm helps to allocate traffic rates to links in a network routing problem where the links have maximum capacities and ensures that no user will receive a traffic rate of 0, while maximizing the flow in the network at the same time. We will give it here out of its historical context in a more compact manner, noting that there are at least three different ways to qualify the result of the BFC algorithm. Independently of the BFC algorithm, the underlying idea of largest minimum has been extensively studied by Rawls in economical science [8]. It is the base for some critique on the utilitarian point of view in economics, summarizing the individual allocations by a kind of utility of a meta-individual. Rawls insists on the point that this says nothings about the modality of the distribution itself and might ignore the base case that some individuals, despite an optimal total might be treated in an unfair manner. The idea to maximize the minimum will not do this and thus be a better base for the fair allocation of resources.

The first representation of this kind of fairness is by the leximin relation. To compare two vectors x and y by this relation (written as $x \geq_{lm} y$) to both vectors their (own) sorting permutations are applied and the resulting vectors with sorted components (non-decreasing order) are lexicographically compared. Note that the specific choice of the sorting permutations does not influence the result. The relation is transitive as well as complete, so usually the maximum sets contain only one element (except permutations of the same vector that belong to the symmetric part of the leximin relation). There is no restriction on the range of the domain.

In order to design a corresponding complementary relation, the syntactic approach gives a relation where the maximal value is minimized. Therefore, the sorted vector has to be reversed before comparison (so both x and y become sorted in non-increasing order) and the comparison is the converse of lexicographic order.

It is easy to see that this relation is suitable for linear regression: it would select a line where the largest deviation from any of the data points is the smallest among all feasible regression lines.

The second representation of maxmin fairness will sort the vector x in non-decreasing order as well, but apply the sorting permutation of x to the components of y (and not the sorting permutation of y as for the leximin relation). Then, again, the permuted vectors are compared by lexicographic order. This is the original definition of maxmin fairness given in [6] (the definition there is different but it can be seen that both are the same relations). Maxmin fairness is cycle-free and implied by Pareto-dominance. Its main difference is that it is kind-of "individualizing" the least component: if we take one component index i and select all elements of the domain where x_i happen to be the smallest value, then the element where this is the largest value is a maximal element of maxmin fairness relation. Simplified speaking, each index has its own maxmin case and no further distinction is done (among them, we also find the leximin maximal element).

The definition of a complement to maxmin fairness can be done like in the leximin case. However, for the regression problem we might find it not so useful to have a regression line for each data point and prefer to use the leximin relation instead.

The third representation is as maximal value of an OWA operator whose weights are exponentially falling [9] as for example $w_i = 2^{(n-i)}$. It gives very large weights to small values (bad allocations) and low weight to large values (allocations to the rich). It demonstrates a role of OWA operators to represent "weighted balance" that will be more clarified in the next subsection.

To summarize, the suitable way to introduce the aspect of least maximum deviation in linear regression is the leximin relation over the domain of feasible regression vectors (i.e. pairs of values (a, b)) and selection of a regression vector by the maximum set of this relation. We will call it leximax in the following.

C. Majorants

Majority of vector x to y (sometimes also called just vector inequality) is based on the majorants of x and y: if the majorant of x Pareto-dominates the majorant of y then x is a *majorant* of y (written as $x \geq_{mj} y$). Initially this relation has been studied in mathematics, but its relevance for economics became soon apparent by e.g.. the Lorentz-curve and related Gini index [10] representing the distribution of wealth in a society. The various properties of majorants are studied in the seminal book [11]. It should be noted that a common assumption of majorants is that the total of x and y is the same, a condition which can be alleviated if taking the general optimization point of view (and not to compare two different ways of distributing *the same total* among a number of individuals).

With regard to linear regression, majorants have a weak point. The weak point is that its maximum sets can be large. One of the major properties of the majorant is that it summarizes all ways of maximizing related OWA operators. it means that if we fix a weight vector and seek the vector x whose OWA with these weights is maximal it will belong to the maximum set of the majorant relation (this is similar to the fact that maximizing a weighted average will give a single element of the Pareto front).

On the other hand, specifying a syntactic complement of the majorant relation is straightforward (we may call it *minorants*): compute the minority vector as vector of partial sums of the vector sorted by non-increasing components, and compare by the complement of Pareto dominance.

However, as we have noted, we still would have to select among the various maximal elements of this relation, and this happens to be the same as looking for suitable weights of an OWA operator that should have a *minimal* value. This is the approach that will be followed here.

D. Proportional fairness

Where the maxmin approach sometimes gives a too strong promotion of the weakest element, proportional fairness seeks a compensation of relative losses and wins. It was originally introduced in [12] and is defined as follows:

Definition 6: Given two vectors x, y from R^n with positive components. Then it is said that x is proportionally more fair than y if and only if

$$\sum_{i=1}^{n} \frac{y_i - x_i}{x_i} \leq 0 \tag{3}$$

This definitions ensures that if considering a change from allocation x to y relative gains are not achieved by stronger losses. The relation (it is like maxmin fairness not always transitive, not complete, but cycle-free and implied by Pareto dominance) implies a larger value of the total logarithmic utility, thus it is often simplified by a comparison

$$\sum_{i=1}^{n} \log x_i \geq \sum_{i=1}^{n} \log y_i \tag{4}$$

With regard to a minimum version, it appears not so easy to come up with a suitable definition (since it includes an arithmetic expression). The main problem is the appearance of division by 0 cases. This holds especially for the linear regression problem, where components can easily be 0 and would basically select lines that pass through data points.

In order to solve this problem, we propose a combination of majorants and proportional fairness to be used as a fairness relation. The relation *proportional minority fairness* (PMF) between x and y then is defined as follows:

1) Calculate the minority vectors of x and y.
2) Compare these minority vectors by the converse of proportional fairness.

The definition would only be invalid in case all data points are exactly located on a single line. Only then, the first component of the minority vector of deviation vector x, which is its largest component, can be 0. However, in this case we do not need to consider a special regression problem as the connecting line will well serve all goals of a linear regression.

In addition to a guaranteed computability, in a convex domain (as the linear regression is) the maximum sets will usually contain only one element. The disadvantage here is that the selection of maximal elements has power two complexity (it needs to consider all pairs of the domain). If sampling a range of (a, b)-parameter pairs to approximate the maximal line, even if its usually not considered intractable, the computational effort can grow rapidly. For example, sampling both, a and b with 1000 values each gives a domain of 1 Million pairs and this includes 10^{12} pairwise comparisons which is a huge computational effort.

However, simplification is possible by taking advantage of the fact that the proposed PMF relation implies a larger relation between the product of components of the minorant vector. Then, the maximal element can be computed by a single pass through the domain as well.

III. Experiments

In the following, we report on a first experiment to study the various ways of linear regression that were introduced in the foregoing section and to learn about their strong and weak points. Moreover, we want to gain some evidence if the fairness aspect is of relevance for the cognitive task of line adjustment or not.

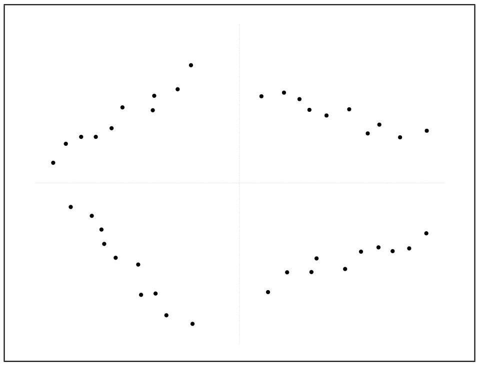

Fig. 1. Sheet with four panels a 10 dots that was used in the experiment.

A. Material and method

The first experiment here was performed among a group of students. Four dot patterns, each composed of ten dots, were presented on a sheet of paper (see Fig. 1). The task was to draw a line by pencil and ruler that is in a "good" balanced position to the points. In fact, all four dot patterns are rotated versions of the same dot pattern.

The students were all familiar with the textbook approach to linear regression, so it was also told to

1) Not necessarily try to "guess" the true regression line but follow their subjective feeling of a good fitting line, representing the linear spread of the dots.
2) Not to just select two points and connect them.

The lines were all drawn in the order upper left panel - upper right panel - lower left panel - lower right panel and without any pause between the drawing of the four lines. In total, seven students incl. one instructor of both gender participated in the experiment. While doing no special questionnaire about their experience, most subjects reported that the task appeared more easy to them before than it was experienced - in fact some difficulties to decide for the "best line" was reported.

B. First evaluation: linear regression

The next step was to measure the positions of points and lines and make elementary statistics. For simplification, i.e. to avoid negative values and to allow direct comparison between all four cases, all measures where mapped into the positive quadrant. The mapping, in terms of the location on the test sheet, where done by measuring both x-and y-coordinates from the center to the outside. The results, including the parameters of the linear regression line, are shown in Table I and a visualization is given in Fig. 2, with the regression line drawn in red (or lighter gray in case of non-color print).

In this figure it can be seen that for three panels the regression line appears more on the border of the distribution of human-selected lines, thus they are a possible but rather unlikely human choice (it should be noted that in 3 of 4 cases, the line most closest to the linear regression line was drawn by the same subject). But for the lower-left panel, the regression line seems to represent the average of the lines as drawn by the experiment subjects. The difference of this

TABLE I

RESULTS FOR AVERAGES (AV) OF SLOPE AND BIAS FOR USER DRAWN REGRESSION LINES IN COMPARISON TO THE LINEAR REGRESSION RESULT (REG) AND FAIR LINEAR REGRESSION FOR LEXIMAX RELATION (LEXIMAX) AND PROPORTIONAL MINORITIES RELATION (PMIN).

	bias	slope
upper-left panel		
av	8.03622	-0.650765
reg	7.58274	-0.59619
leximax	7.524	-0.606
pmin	7.6	-0.6
upper-right panel		
av	5.4625	-0.291071
reg	5.34656	-0.273577
leximax	5.291	-0.279
pmin	5.38	-0.28
lower-left panel		
av	11.0955	-1.06708
reg	11.3806	-1.08912
leximax	11.394	-1.072
pmin	11.36	-1.08
lower-right panel		
av	6.35714	-0.342857
reg	6.09127	-0.303139
leximax	6.062	-0.309
pmin	6.04	-0.3

TABLE II

THE 22 MAXIMAL ELEMENTS FOR THE MINORITIES RELATION FOR THE UPPER-LEFT PANEL.

slope	bias
-0.66	7.96
-0.66	8.0
-0.64	7.76
-0.64	7.8
-0.64	7.84
-0.64	7.88
-0.64	7.92
-0.62	7.6
-0.62	7.64
-0.62	7.68
-0.62	7.72
-0.62	7.76
-0.62	7.8
-0.6	7.52
-0.6	7.56
-0.6	7.6
-0.6	7.64
-0.6	7.68
-0.58	7.44
-0.58	7.48
-0.58	7.52
-0.56	7.36

panel to the other panels is that here the dot pattern is most steeply falling.

We see this as evidence that in fact the human solves this cognitive task by focussing on the "landmark" dots — thus basically taking the distance in y-direction more strongly into account than constructing a hypothetical virtual point in mind that is located nearby the given point and whose position might be closest to a candidate line. In case of the lower-left panel, the closest points to the regression line will be less offset from the x-positions of the given dots and thus the linear regression line more likely. With this reasoning, in the following we choose the y-distance only as a measure for deviation of a line to the given dots. In the next experiments, we want to investigate whether the fair linear regression lines could give a better representations of the human choice of a balanced line than the linear regression.

C. Second evaluation: fair linear regression

With regard to select maximum sets of the fairness relations that were introduced and selected in the foregoing section, we performed evaluations as follows:

- For maxmin fairness, we computed parameters of a regression line such that the largest y-distance of a point had the smallest value. The y-distances here serve as deviation vector. The parameters where selected by subsampling a parameter range of width 1 around the linear regression values, subdivided into 1000 intervals.
- For proportional minority, we took a smaller subsampling of 100 intervals for the two parameters and computed the maximum set of the relation by exhaustive pairwise comparison (i.e. 10000 comparisons). Since we have to use the asymmetric part of the relation to confirm maximality, in each case where x (i.e. a specific choice of parameters) was in relation to y it had to be tested that not y happen to be in relation to x as well.
- For minority fairness (the complement relation to majorants) same settings were used as in the case before, and the maximum set was exhaustively sampled.

Table I also shows the results for these cases. In all cases the adapted lines have parameters more close to the linear regression line than to the user average selection. This is the basic observation that can be done, and it seems that there is no specific evidence for a user selection based on minimizing the maximal deviation, or balancing the deviations.

However, minority fairness gives a different picture. In Table II the set of 22 maximal elements of the minority relation for the upper-left panel is listed and shows that indeed some of them cover the average user selection better than linear regression (especially the second pair -0.66 and 8.0 is very close to the user average). But the fact that the number of candidates is rather large makes this approach not practicable. So we have to consider that each of these maximal elements is related to the minimum value of some OWA with corr. weights. In the third evaluation, we want to explore the adaptation of OWA weights.

D. Third evaluation: adaptation of OWA weights

The result of the foregoing evaluation was an indication for flexibility of the OWA operator to represent the user selection of line parameters more closely than linear regression. We want to compare this with the result of using weights in the linear regression. This way it can be seen if the position of a dot has stronger influence on the human line drawing or the distribution of distances. For adapting the 10 weights in both cases, we used Differential Evolution (DE).

The choice for DE is based on recent good experiences with DE for a broad range of optimization tasks, as well as ease of implementation, direct representation of real values, and low number of parameters. Since it is a well known procedure there is no need to repeat the algorithm here. Most relevant here is the circumstance that both adaptations (for OWA weights, and for weighted linear regression) are performed under the same settings of DE parameters, and thus also the same number of function evaluations.

For the case of OWA weights, the fitness computation for a single individual s is performed as follows:

1) The individual p is encoding a weight vector of an OWA.
2) The parameter range of size 1 for slope and bias each is subdivided into 1000 intervals around the linear regression value, and for these 1 Million cases the parameters a_{min}, b_{min} are found that minimize the OWA of y-distances using p as weights.
3) The fitness (here its a minimization task) is given by

$$F = 10|a_{min} - a_{av}| + |b_{min} - b_{av}| \qquad (5)$$

where a_{av}, b_{av} are the average slope and bias of the human-selected lines, and the factor 10 is introduced for compensating the different numerical ranges of a and b.

The corresponding fitness for the weighted linear regression is basically the same procedure, except that in case of the OWA the weighted average is used. It should be noted here that also for the weighted linear regression, we only use y-distances at the data points and not the shortest distance to the line.

The settings for DE were chosen by values that gave rather fast convergence: population size is 20, the weight values (i.e. components of the individual vector) were clamped between 0 and 1 including 0 and 1, the injection rate was set to 0.5 and the parameter F to 1. DE was performed for 100 steps. For a wider range of DE parameters there were actually no different results, but a larger number of generations was needed to achieve same performance.

Table III gives some results. A notable feature of using DE in this context was that the fitness values after 100 generations, per panel, with very few exceptions, where always the same in both cases, OWA and weighted linear regression. For illustration, a few weights vectors are shown as well. The following observations can be made:

- In all cases the OWA comes more close to the user average than the weighted linear regression. For the upper-right panel it is a very large difference, for the lower-right panel more a close match. This is a clear indication that the OWA can be much better adjusted to the human way of balancing lines then the weighted linear regression. Thus, there is some indication that human observers take the distribution of distances into account.
- While the best fitness value appears to be the same, the weights themselves vary a lot. It shows that same results can be achieved within a wide range of weights, for OWA as well as weighted linear regression. In case

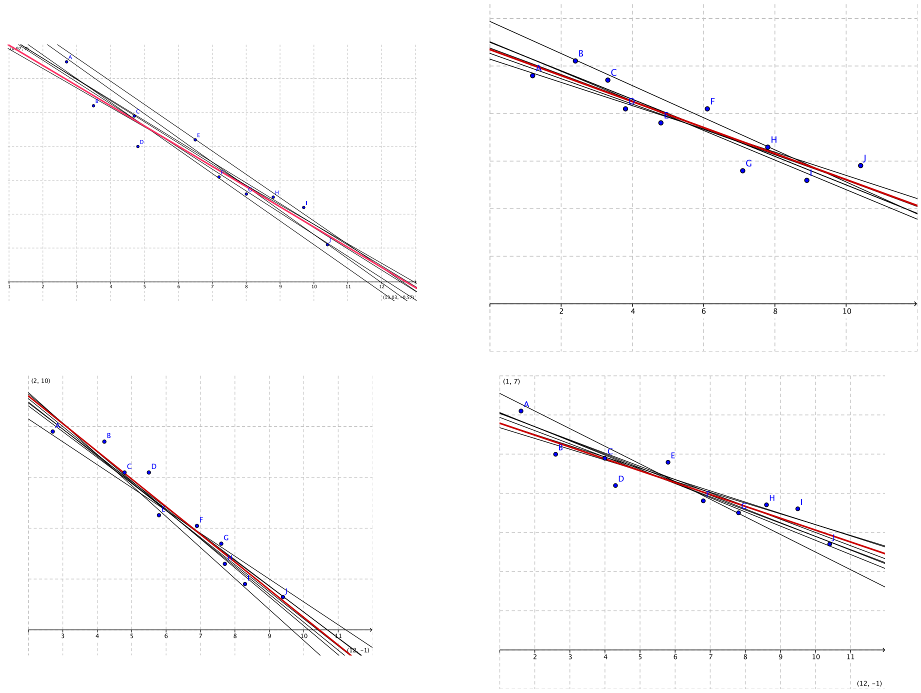

Fig. 2. User-drawn regression lines (in black) and linear regression lines (in red) for the four panels of the experiment sheet. Note that the dot positions and lines were mapped into the positive quadrant, and also that the four subfigures have different numerical scales in x-direction (horizontal axis) and y-direction (vertical axis) for better visualization.

TABLE III

SELECTED BEST RESULTS FOR OWA (OWA) AND WEIGHTED LINEAR REGRESSION (WLR) ADJUSTED TO THE USER AVERAGE REGRESSION LINE BY USING DIFFERENTIAL EVOLUTION.

	example weights	fitness
upper-left panel		
owa	(0.291195 0 1 0.171037 0.143524 1.0 1 0.25157 0.722466 0)	0.0114306
wlr	(1 0.417243 1 1 0.149854 0.0239066 1 0.699275 1 1)	0.128569
upper-right panel		
owa	(0 0.043132 0.245781 0.168147 0.496377 0.616693 1 1.0 0.569346 0.924157)	0.0182095
wlr	(0 0.754094 1.0 0 1.0 0 0.644548 0.306872 0.114402 0.0)	4.04387
lower-left panel		
owa	(0.0 0 0.653261 0.209838 0.531429 0.0 1 0.896185 0 0.538717)	0.0337007
wlr	(0.564098 0.742391 1 0.129688 0.767997 0 0 0.04389 0.114336 0.918784)	0.235301
lower-right panel		
owa	(0.257143 0.974894 1 0 1 1.0 0 0.742932 1 0.190612)	0.0557098
wlr	(1 0.203465 1 1 1.0 0.295419 0.637501 0.477884 0 0.749796)	0.0885701

of OWA it should be also taken into account that the weights for the smaller deviations (which can become very close to 0 or even 0) do not have a strong numerical influence on the adaptation, so it is not expected to have stable values here.

- In case of the upper-left panel it is interesting to note that the adjusted OWA weights in nearly all cases have a rightmost weight (for the largest deviation) that is smaller than the second one from the left (the weight for the second largest deviation). However, this is not the case for the other panels. In this case one might consider the perception of an outlier dot that is ignored in the line drawing task and producing a largest deviation. Then, the user is focusing on the next point.
- The weights adapted for the weighted regression show no general pattern. Often, 0 and 1 weights appear, but in varying positions. There can be the same best fitness achieved with canceling the left- or rightmost point, but

also with a strongest weight of 1. It shows that there is no positional preference of the user to place the line, e.g. by using the left- or rightmost as "support points."
- It is not explicitly given by results, but the continuation of the DE for 500 generations did not change the results.

In summary, where the direct application of a fairness relation for regression gave no clear difference to the linear regression, especially no better adaptation to the user choice, the adaptation of weights was clearly much better in case of the OWA operator. We got some evidence that user take distribution aspects of deviation of a line from given points into account.

E. Limitations

The comparison of linear regression with different approaches to fair linear regression was only possible by proposing a concept of using maximality of general relations for this task. Hopefully, the foregoing analysis has demonstrated the practical application of this concept. However, we are also aware that the drawn conclusions are limited in their scope, due to a number of reasons:

- First at all, the number of subjects was rather small, and constant problems were used in all cases. It needs a larger base and also a procedure for automatically creating a number of reasonable tasks to improve the confidence into the obtained results.
- Contrary to linear regression, the general problem does not have an analytic solution. There is some effort needed to find the maximal elements of the used relations, with additional parameter settings and granulation effects.
- Even if in nearly all cases the same best fitness values were obtained, there is no guarantee that these are the global best values. This is a known fact for any metaheuristic algorithm like DE.
- The way of representing the selection of all subjects was done by averaging slope and bias. It might be questioned if not another way of fusing these parameters should be used, as the same arguments on linear regression and the related MSE minimization might also apply here.
- Only DE was used. By its nature, DE gives an easy framework for such adaptation tasks, but a comparison with other metaheuristic algorithms might not give better results but insight into the stability of the results.
- The constellation of dots seems to have influence on how the human solves the cognitive task of line balancing. Probably it needs many more specific experiments to find out about human preferences and also perception of outliers.

Needless to say, these limitations are within the experimental design and do not demonstrate (except a possible larger computational effort) any essential drawback of the proposed concept of fair linear regression.

IV. Summary

We started with a carefully performed critique on the use of linear regression. We feel that in tasks where human cognition of visual scenes, or in other regard allocation and resource sharing tasks are considered, the standard approach of linear regression to minimize the MSE might not be appropriate. This is the motivation to have alternatives at hand. By using fairness relations that incorporate aspects of the distribution of items, a suitable set of relations was identified that can be used in an alternative way to standard linear regression: by finding maximum elements, i.e. elements of the asymmetric part of a relation to which no other element is in relation. For using this concept it was necessary to provide minimum versions of fairness relations.

The approach was demonstrated for the cognitive task of human-placed regression lines for a specific dot pattern. While linear regression comes close to the human-drawn lines, the precision can be even more improved by evolutionary adapted OWA operator weights. Here, an evolutionary approach is suitable since the number of weights is equal to the number of dots, and in general there are no analytic ways to solve the corr. extremity problems. In an experimental setup a related analysis showed that using OWA can give the best representation of the human selection of a fitting line, especially notably better than a weighted linear regression.

Acknowledgments

The authors thanks the students of "Kaori Lab" at the Kyushu Institute of Technology for their generous support of the experiments reported in this study.

References

[1] J. Y. L. Boudec. Rate adaptation, congestion control and fairness: a tutorial. [Online]. Available: http://ica1www.epfl.ch/PS$_{f}iles/$-LEB3132*.pdf*

[2] M. Köppen, "Relational optimization and its application: From bottleneck flow control to wireless channel allocation," *INFORMATICA*, vol. 24, no. 3, pp. 413–433, 2013.

[3] D. Kahneman and A. Tversky, "Choices, values, and frames." *American psychologist*, vol. 39, no. 4, pp. 341–350, 1984.

[4] M. Köppen, K. Yoshida, and K. Ohnishi, "A gratis theorem for relational optimization," in *Proc. 11th International Conference on Hybrid Intelligent Systems (HIS 2011)*, Melaka, Malaysia, December 2011, pp. 674–679.

[5] K. Suzumura, *Rational Choice, Collective Decisions, and Social Welfare*. Cambridge University Press, 2009.

[6] D. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs, N: Prentice Hall, 1992.

[7] J. Jaffe, "Bottleneck flow control," *IEEE Trans. Commun.*, vol. COM-29, July 1981.

[8] J. Rawls, *Justice as fairness: A restatement*. Harvard University Press, 2001.

[9] M. Köppen, J. Okamoto, and A. Honda, "Fuzzy fusion fairness relations for the evaluation of user preference," in *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011), Proceedings*, Taipei, Taiwan, June 2011, pp. 2566 – 2574.

[10] C. Gini, "Concentration and dependency ratios (english translation)," *Rivista di Politica Economica*, vol. 87, pp. 769–789, 1997.

[11] A. W. Marshall, I. Olkin, and B. Arnold, *Inequalities*. Springer, 2011.

[12] F. Kelly, "Charging and rate control for elastic traffic," *Eur. Trans. Telecomm.*, vol. 8, pp. 33–37, Jan./Feb. 1997.