# Distance equalizers

Thomas Kämpke
Forschungsinstitut für Anwendungsorientierte Wissensverarbeitung/n FAW/n
Lise-Meitner-Str. 9, 89081 Ulm, Germany
kaempke@faw-neu-ulm.de

**Abstract**

Distance equalizers are introduced as empirical measures of central tendency that make distances to univariate data as similar as possible. These measures are made precise by means of various so-called fluctuation functions which account for distances in different ways. Distance equalizers differ from the mean and the median. Also, distance equalizers relate to dispersion measures like the median of absolute deviations and allow to define new dispersion measures. Algorithms as well as a closed form solutions for special cases are given. Computations require to perform multiextemal function minimization of several kinds.

Distance equalization is shown to enable cluster analysis with a special neighbouhood notion. Cluster computations are reduced to computations of shortest paths with prescribed number of intermediate nodes in weighted directed graphs.

Key words: central tendency, dispersion, mean value, median, robust statistics, shortest path computations.

## 1 Problem

Common empirical measures of central tendency like the mean value and the median of a finite set of data express some kind of average that represents all data. The dispersion around such a central value can be expressed by the standard deviation, the skewness, the range etc. What matters here is distance between each data point and a suitable reference value such that these distances vary only little. A simple example is the set of the four data 1.9, 2.0, 2.1 and 12.0. The value 7.0 has about equal distance to all four data but this so-called distance equalizer lies relatively far from the median 2.05 and from the mean value 4.5 of the data. Also, since the harmonic mean is always smaller than the geometric mean which, in turn, is always smaller than the arithmetic mean, these two other means are further apart from the distance equalizer than the usual mean.

One motivation for this investigation is the task of clustering data by quality in terms of their fluctuation around some unknown reference value or around some known function. The degree of fluctuation may depend on certain regions and the task is to identify these regions. Crucial to this task is not to minimize fluctuations around some reference value or around a given function but to some average fluctuation around the reference value or the function. Intuitively, fluctuations should be equal rather than small. This allows to cluster data into classes of about equal fluctuation. A feature of an optimal fluctuation clustering is that data with large and even largest distance may fall into the same class.

Distance equalizers will be formulated as global minimizing points of so-called fluctuation functions. These aggregate so-called constituing functions either by summation or by forming medians. The constituing functions themselves are obtained either by summation or by forming medians. All distance equalizers can be considered as measures of central tendency while certain values of some of the fluctuation functions turn out to be popular dispersion measures. In addition, the global minimum value of any fluctuation function can be considered as a (new) dispersion measure.

All but two types of distance equalizers cannot be computed by closed formulas. Enumerations of a finite number of candidate locations will then lead to the distance equalizers in many cases. In most of theses cases, the candidates to be inspected are zeros of certain constituing functions and the number of these zeros is bounded linearly in the number of the given data. In the exceptional case of the constituing functions and their aggregation both being formed by medians, search among the zeros is insufficient and must be extended to the intersection points of the constituing functions. The number of the latter is no more linearly bounded in the number of the data. In all non-closed formula cases, it is required to compute zeros of piecewise linear, quadratic and other functions and to evaluate the fluctuation functions there.

Fluctuation functions allow to cluster data into classes such that fluctuations within each class become small while fluctuations between classes are ignored. Algorithms for fluctuation clusterings are more complicated than for mere equalizer computations. Cluster computations will be related to computing shortest paths with a prescribed number of intermediate nodes in particular directed graphs. A modification of the Dijkstra algorithm will be given that efficiently solves this problem. Computing the arc weights will then turn out to be the algorithmic bottleneck since each weight requires to find a distance equalizer over a subset of the given data.

Related work on central measures of tendency mostly considers the classical quantities mean, mode and median. These have been analyzed extensiviely from methodological and applicational perspectives and have also been modified such as to the trimmed mean. Relatives of the usual or arithmetic mean are the geometric and the harmonic mean which are used less frequently since they are more difficult to motivate. In addition, higher central moments, quantiles and some combinations of quantiles are occasionally considered as central measures as well. M-estimators intend to combine features of the mean and the median by means of weighted sums of the data.

Potentially more subtle measures of central tendency, dispersion, skewness etc. have been constructed from moments of order statistics in [Ho] but for centrality these do not reach beyond the mean value. Continuous distributions which do not have finite expectations are assigned central measures in the recent work [Fa] and median estimators for certain symmetric distributions without finite expectations are given in [MOPP]. Central measures over special geometric arrangements are considered in [Si].

The remainder of this work is organized as follows. Various notions of distance equalizers for real-valued data and some of their properties are introduced in section 2. Section 3 proposes closed formulas for the computation of distance equalizers for the few cases in which this is possible and algorithms in all other cases. Cluster approaches based on distance equalizers are introduced in section 4 and cluster computations are sketched in section 5.

# 2 Approach

## 2.1 Concept and definitions

The second centered moment of a finite set of real-valued data is, as usual, denoted as $var(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - x)^2$ with standard deviation $std(x) = \sqrt{var(x)}$. It is well known and easily verified that the second centered moment is minimized by the mean value of the data which means that

$$argmin_{x\in I\!R}\ \frac{1}{n}\sum_{i=1}^{n}(x_i - x)^2 = \frac{1}{n}\sum_{i=1}^{n}x_i = \bar{x}.$$

Instead of minimizing all distances between the data and a candidate value, here, so-called fluctuation functions and fluctuation measures will be introduced that intend to equalize all the distances between the data and a candidate value. Thus, the distances should differ as least as possible from an average distance rather than be as small as possible.

Fluctuation functions differ in the particular way in which (1) individual distances are computed, (2) their average is computed and (3) in the way all distances and their average are aggregated. In particular, taking squares will be generalized and summation as well as taking medians are the essential operations for averaging and aggregation. For reasons that become obvious later, fluctuation functions can also be denoted as empirical dispersion functions.

Functions of many types are suitable, in principle, and an important collection is that of the real-valued Lipschitz fluctuation functions

$$L_p(x) = \sum_{i=1}^{n} \Big| |x_i - x|^p - \frac{1}{n} \sum_{j=1}^{n} |x_j - x|^p \Big|^p.$$

Clearly, the rationale of these functions is that individual fluctuations around some value should come as close as possible to the average fluctuations around that value with both average and proximities to this average varying with some shape parameter $p \in [1, \infty)$. Average fluctuations are expressed by the mean deviation functions $f_p(x) = \frac{1}{n} \sum_{j=1}^{n} |x_j - x|^p$. Mean deviation functions generalize the second centered moment since $f_2(x) = var(x)$. The case $p = \infty$ is covered by the fluctuation function

$$L_\infty(x) = \max_{i=1,\ldots,n} \Big| |x_i - x| - \max_{j=1,\ldots,n} |x_j - x| \Big|.$$

Depending on the shape parameter and the data, a Lipschitz fluctuation function may be multiextremal which means that it may posess several local minima and even several global minima. A fluctuation function may hare cusps, regions of constant values and it is constantly zero in the trivial case of all data being equal. Any global minimizer is denoted as the $L_p$-distance equalizer of the data or, shortly, the $L_p$-equalizer. A slightly different fluctuation function for finite shape parameters is obtained from replacing the mean deviation function by the standard deviation

$$K_p(x) = \sum_{i=1}^{n} \Big| |x_i - x|^p - std(x)^p \Big|^p.$$

These modifed fluctuation functions are more curved than the Lipschitz fluctuation function and are thus also denoted as curved Lipschitz fluctuation functions. Global minimizers of the curved Lipschitz fluctuation functions are denoted as $K_p$-equalizers. Nothing is changed for shape parameter two which means $K_2(x) = L_2(x)$ for all $x$, but almost every function value changes for shape parameter one which means that the modified and the original Lipschitz fuctuation functions are different for almost all arguments. Obviously, different shape parameters, the mean deviation function and the standard deviation can be combined in more ways to result in additional fluctuation functions of the Lipschitz type.

Yet another variation of the Lipschitz fluctuation function for shape parameter one is obtained from replacing the mean deviation function by the centered median function. This results in the median fluctuation function

$$M_1(x) = \sum_{i=1}^{n} \Big| |x_i - x| - median(|x_1 - x|, \ldots, |x_n - x|) \Big|.$$

The motivation herefore is to center the fluctuations around an average that minimizes the fluctuation function for every given value of $x$. This minimizer of a sum of absolute differences is known to be the median which is formally denoted as $argmin_w \sum_{i=1}^{n} |w_i - w| = median\{w_1, \ldots, w_n\}$. Applying the median not only to the distance average but to the aggregation as well results in the double median fluctuation function

$$MM_1(x) = median_{i=1,\ldots,n} \Big\{ \Big| |x_i - x| - median(|x_1 - x|, \ldots, |x_n - x|) \Big| \Big\}.$$

Finally a distance equalizing objective is mentioned that seemingly does not compare individual distances to some kind of an average distance. This leads to the range equalizer which simply minimizes the range fluctuation function

$$R(x) = \max\{|x_1 - x|, \ldots, |x_n - x|\} - \min\{|x_1 - x|, \ldots, |x_n - x|\}.$$

It can be seen that the range equalizer is equivalent to the Lipschitz equalizer with shape parameter infinity since both fluctuation functions are equal. Though data are univariate, minimizing their fluctuation amounts to finding a circle that best fits the data according to different measures of fit, see figure 1. The distance equalizer is the center of the circle and the average fluctuation around the center indicates the radius. This allows to consider multivariate problem versions such as circular regression in two dimensions [Sh] but, here, the focus is on univariate problems. Anyway, it should be noted that fluctuation function minimization is not completely identical to, for example, the standard notion of least squares distances and least absolute distances as used in circle fitting.

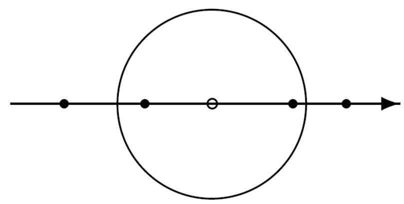

Figure 1: Four data on the real line and fitting circle with center on the real line.

Sample values of various equalizers are given in the following table indicating no consistent monotonicity of the equalizer as function of the shape parameter. All computations and some of drawings were done in Scilab [Sci].

| Shape parameter $p$ | $L_p$-equalizer | $K_p$-equalizer | $M_1$-equalizer | $MM_1$-equalizer | $R$-equalizer |
|---|---|---|---|---|---|
| 1.0 | 8.450 ... 8.950 | 15.276 | 7.350 | 6.750 ... 7.100 | – |
| 1.2 | 15.610 | 15.231 | – | – | – |
| 1.4 | 14.513 | 14.471 | – | – | – |
| 1.6 | 13.662 | 13.849 | – | – | – |
| 1.8 | 13.401 | 13.420 | – | – | – |
| 2.0 | 13.294 | 13.294 | – | – | – |
| 2.2 | 13.278 | 13.288 | – | – | – |
| 2.4 | 13.293 | 13.321 | – | – | – |
| 2.6 | 13.318 | 13.363 | – | – | – |
| 2.8 | 13.343 | 13.404 | – | – | – |
| 3.0 | 13.368 | 13.438 | – | – | – |
| 4.0 | 13.452 | 13.513 | – | – | – |
| 5.0 | 13.489 | 13.518 | – | – | – |
| 6.0 | 13.504 | 13.514 | – | – | – |
| 7.0 | 13.509 | 13.511 | – | – | – |
| 8.0 | 13.509 | 13.508 | – | – | – |
| 9.0 | 13.508 | 13.506 | – | – | – |
| 10.0 | 13.507 | 13.504 | – | – | – |
| $\infty$ | 13.500 ... 18.600 | – | – | – | 13.500 ... 18.600 |
| mean value | | | 9.617 | | |
| median | | | 8.000 | | |

Table 1. Distance equalizers for the $n = 6$ data $2, 2.5, 5, 11, 12.2, 25$. All except the $L_1$-, $L_\infty$- and the $MM_1$-equalizer are unique and all equalizer values are rounded to three decimal digits.

Possible multiextremality of a Lipschitz fluctuation function is indicated in figure 2. Whenever the global minimizers form an interval and a single value is opted for, the interval midpoint is denoted as equalizer. When a single value is not opted for, any global minimizer is a valid distance equalizer.

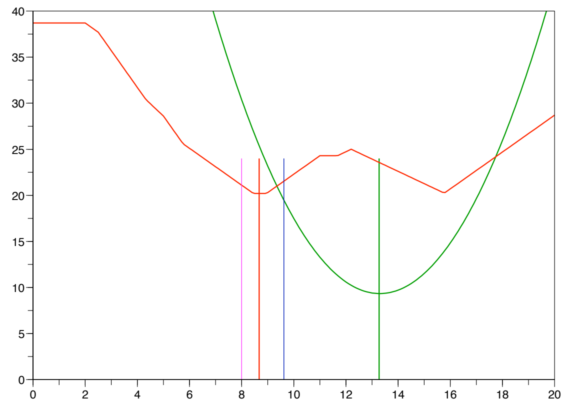


Figure 2: Graphs of the $L_1$-fluctuation function with multiple minima and of the $L_2$-fluctuation function with single minimum for the data from table 1. Four measures of central tendency, namely median, $L_1$-equalizer, mean value and the $L_2$-equalizer are indicated by vertical bars (from left to right). Values of the $L_2$-fluctuation function are divided by 2000 so that they become similar to other function values; the position of the $L_2$-equalizer thus remains unaffected.

Two modified Lipschitz fluctuation functions and their equalizers are sketched in figure 3. Not only the functions, but also, their equalizers can be seen to differ sharply for shape parameter one. The $M_1$-equalizer and the $MM_1$-equalizer are the smallest of all equalizers for the data from table 1 as shown in figure 4.

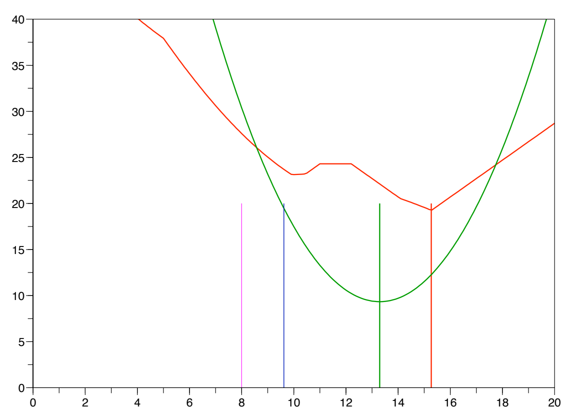


Figure 3: Graphs of the $K_2$-fluctuation function with single minimum − the same as in figure 2 − and of the $K_1$-fluctuation function with multiple minima for the data from table 1. Four measures of central tendency, namely median, mean value, $K_2$-equalizer (= $L_2$-equalizer) and $K_1$-equalizer are indicated by vertical bars (from left to right). Again, the values of the fluctuation function $K_2$ are divided by 2000.

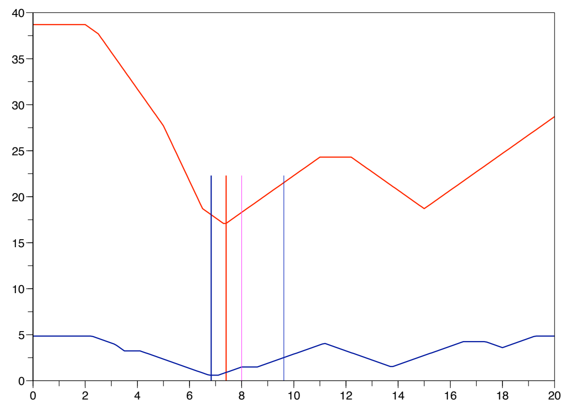


Figure 4: Graph of the $M_1$-fluctuation function (upper curve) and $MM_1$-fluctuation function (lower curve) for the data from table 1. Four measures of central tendency, namely $MM_1$-equalizer, $M_1$-equalizer, median and mean value are indicated by vertical bars (from left to right).

## 2.2 Properties

Global minimizers of fluctation functions generally do not coincide with any minimizer of the corresponding mean deviation function, standard deviation function or centered median function. All mean deviation functions $f_p(x)$ and the standard deviation function are always convex while the centered median function need not be convex.

For shape parameters between one and two, the global minimizer of the mean deviation function lies between the mean value and the median of the data. This interpolates between the case $p = 1$ that the sum of absolute distances is minimized by the median and the case $p = 2$ that the sum of squared distances is minimized by the mean value. Equalizers themselves need not lie between mean value and median. The $L_2$-equalizer not lying there is shown in figure 2 and an example of the $L_1$-equalizer not lying there is of smallest possible size, namely the three data $x_1 = 2$, $x_2 = 4$, $x_3 = 6$. Their mean value, median, and $L_2$-equalizer all are four, but the $L_1$-equalizer, which is twofold here, has the values 3.5 and 4.5.

The same applies to the $K_1$-equalizer which has the twofold values 3.667 and 4.333 for the foregoing three data. It can be stated as a rule of thumb that, in case of a "symmetric" data set of odd size, equalizers for shape parameter one tend to break away from symmetry while they tend towards symmetry for shape parameter two. The double median equalizer has the three values 3,4,5 for the foregoing data as the double median fluctuation function always has three global minimizer of value zero for three data.

### 2.2.1 Scaling

When data are scaled by a common factor, equalizers and optimal values of fluctuation functions undergo the same scaling. But when the data undergoe distorted scaling so that distances between some data are preserved, the effect is inhomogenous across fluctation functions. Optimal values of Lipschitz fluctuation functions with shape parameter one may remain constant while others change as shown in table 2.

| Shape parameter | Optimal value $\min_{x\in I\!R} L_p(x)$ | |
|---|---|---|
| | $x_1 = 0, x_2 = 9, x_3 = 10$ | $x_1 = 0, x_2 = 99, x_3 = 100$ |
| p=1 | 1 | 1 |
| p=2 | 44.5055 | 4949.50 |

Table 2. Minimum values of fluctation functions for three data points with constant difference between second and third value.

#### 2.2.2 Limiting values

The single and double median fluctuation functions and the original Lipschitz fluctuation function for shape parameter one all are bounded since they are constant outside the data range

$$
\begin{aligned}
L_1(x) &= \sum_{i=1}^{n} |x_i - \bar{x}| \\
M_1(x) &= \sum_{i=1}^{n} |x_i - median\{x_1, \ldots, x_n\}| \\
MM_1(x) &= median_{i=1,\ldots,n} \{|x_i - median\{x_1, \ldots, x_n\}|\},
\end{aligned}
$$

for all $x \geq \max\{x_1, \ldots, x_n\}$ and all $x \leq \min\{x_1, \ldots, x_n\}$. The constant values are considered as empirical dispersion measures. They are denoted, respectively, as absolute deviation from the mean, absolute deviation from the median, and median absolute deviation (from the median) MAD, see, for example [MaBr]. More often, the first two values are divided by $n$ to result in the mean absolute deviations from the mean and the median. It is this feature of attaining certain popular dispersion measures which leads to the alternative denotion of fluctuation functions as empirical dispersion functions.

As a consequence, the empirical dispersion measures can be refined by inserting distance equalizers instead of extreme data into the empirical dispersion functions. For example, the absolute deviation from the mean then becomes the optimal absolute value of the Lipschitz fluctuation function with shape parameter one. Symbolically, these replacements are

$$
\begin{aligned}
\sum_{i=1}^{n} |x_i - \bar{x}| &\rightarrow L_1(x_{L,1}) \\
\sum_{i=1}^{n} |x_i - median\{x_1, \ldots, x_n\}| &\rightarrow M_1(x_{M,1}) \\
median_{i=1,\ldots,n} \{|x_i - median\{x_1, \ldots, x_n\}|\} &\rightarrow MM_1(x_{MM,1}).
\end{aligned}
$$

More systematically, the Lipschitz dispersion measures and the modified Lipschitz dispersion measures for general shape parameters can be defined as $L_p(x_{L,p})$ and $K_p(x_{K,p})$ respectively. This puts some burden on the computations of distance equalizers, see section 3. Computations of empirical dispersion measures involving medians are straightforward since median computations being simple for sorted data carry over to MAD computations. When the data are sorted increasingly and when, for example, the size of the data set is odd at $n = 2\,k + 1$, the median and the median absolute deviation can be expressed as

$$
\begin{aligned}
median(x_1, \ldots, x_n) &= x_{k+1} \\
median_{i=1,\ldots,n} \{|x_i - median\{x_1, \ldots, x_n\}|\} &= \min(x_n - x_{k+1}, x_{k+1} - x_k).
\end{aligned}
$$

The modified Lipschitz function $K_1(x)$ is not piecewise linear though segments may seem to be so. Outside the data range, the function is not constant and may attain its maximum there. Yet, the function is bounded on the real line since it converges, namely to the level of the original Lipschitz fluctuation function

$$\lim_{x\to\infty} K_1(x) = \lim_{x\to-\infty} K_1(x) = \sum_{i=1}^{n} |x_i - \bar{x}|.$$

This result can be obtained from the asymptotic behaviour of the standard deviation $\lim_{x\to\infty} std(x) - |x-\bar{x}| = \lim_{x\to-\infty} std(x) - |x-\bar{x}| = 0$. Global minima of fluctuation functions need only be searched in the data range since Lipschitz fluctuation functions, for example, only attain constant or larger values outside. Though not immediately relevant for distance equalizing, the global maximum of a fluctuation function with shape parameter one may also be attained within the data range.

Multiextremality of the Lipschitz and modified Lipschitz fluctuation functions has only been observed for shape parameters less than two. For shape parameter one the number of local minima can grow linearly in the number of data. Samples with more than two local minima are shown in figure 5.

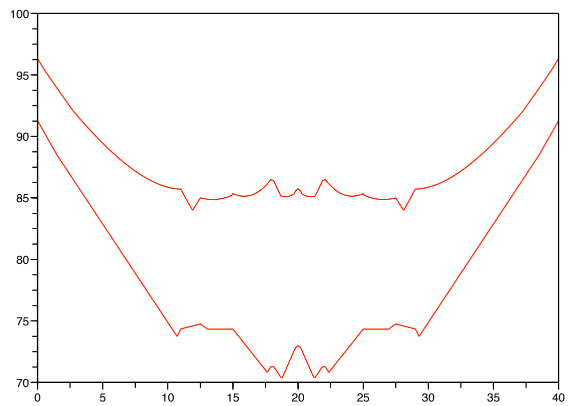


Figure 5: $K_1$-fluctuation function (upper curve) with eight local minima and $L_1$-fluctuation function (lower curve) with six local minima of which two are small intervals that constitute the global minimizer. Both curves refer to the same set of $n = 15$ data that lie symmetric in their range with $mean(x_1, \dots, x_{15}) = median(x_1, \dots, x_{15}) = 20$.

### 2.2.3 Missing monotonicity in the data

Distance equalizers generally do not increase with the data. This is different from the monotonicity behaviour of mean values and medians and it is made precise as follows. When one or more data increase by an arbitrary amount and all other data remain fixed, both their mean value and median increase. Yet for the four data $x_1 = 1$, $x_2 = 2$, $x_3 = 3$, $x_4 = 12$, two of their distance equalizers are $x_{L,1} = 6.995$ and $x_{L,2} = 6.837$. When the third value increases to $x_3 = 11$, the distance equalizers decrease to $x_{L,1} = 6.005$ and $x_{L,2} = 6.5$. The same behaviour is exhibited by all other distance equalizers with the situation for the single and double median equalizers being slightly more intricate because of their quite large regions of global optimality.

Even more, monotonicity may be absent if data increase only moderate so that no value crosses the original distance equalizer. This can be seen from only three data $x_1 = 2$, $x_2 = 3$, $x_3 = 5$. Their $L_1$-

and $L_2$-distance equalizers are both unique at values 3.750 and 3.572 respectively, but when the second value is increased to $x_2 = 3.55$, both distance equalizers decrease to $x_{L,1} = 3.137$ and $x_{L,2} = 3.492$ respectively. The situation for shape parameter one is sketched in figure 6. Originally, the effect of the middle data value on the distance equalizer is an upward push, while it reverses to a downward push by a data increase.

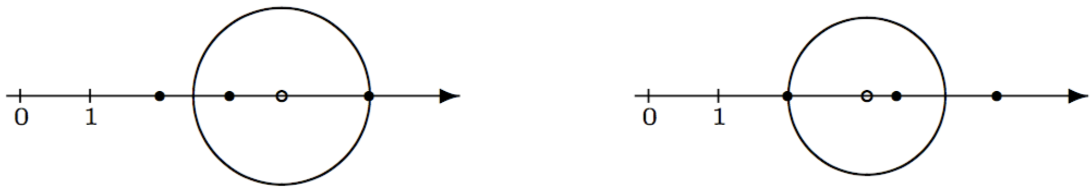


Figure 6: Original situation with two data close to the lower section of the best-fit circle and one data value in opposite proximity (left). The situation is reversed for slightly increased middle data point (right). The best-fit circle shrinks slightly.

### 2.2.4 Convergence in the data

The double median distance equalizer is the only distance equalizer which exhibits a certain robustness in the sense of breakdown points. The breakdown point of an estimator is understood as the proportion of data which can be driven towards inifinity with the estimator remaining bounded [LoRo]. A mean value can be driven beyond any limit by just one data tending to infinity so that a mean value has the smallest possible breakdown point zero and is considered as not being robust.

But the breakdown point of the median equals $\lfloor \frac{n-1}{2} \rfloor / n$; the floor $\lfloor x \rfloor$ denotes the largest integer which is less or eaqual to $x$. Thus, almost 50% of the data can increase beyond any limit leaving the median bounded. This feature is considered as robustness of the median. Even more, the median remains unchanged if up to $\lfloor \frac{n-1}{2} \rfloor$ data above the median tend to infinity and all other data remain fix.

All distance equalizers except the double median distance equalizer have breakdown point zero and the double median distance equalizer itself has the smallest non-trivial breakdown point $1/n$. This means that whenever one of the data grows arbitrarily large in absolute terms, the double median distance equalizer remains bounded. Only in certain cases the equalizer remains even unchanged.

This mild degree of robustness can be derived as follows. In order to avoid too many technicalities, the number of data is assumed to be odd with $n = 2\,k + 1$, the data are pairwise different and increasingly sorted as $x_1 < \ldots < x_n$ and it is the largest value that increases from $x_n$ to $x_n + \mu$ with $\mu > 0$. This increase leaves the median of all data unaffected as $x_{k+1} = \mathit{median}(x_1, \ldots, x_n + \mu)$ and the median of absolute deviations attains the values

$$
\begin{aligned}
&\mathit{median}\{|x_1 - \mathit{median}\{x_1, \ldots, x_n + \mu\}|, \ldots, |x_{n-1} - \mathit{median}\{x_1, \ldots, x_n + \mu\}|, \\
&\qquad\qquad |x_n + \mu - \mathit{median}\{x_1, \ldots, x_n + \mu\}|\} \\
&= \min(x_n + \mu - x_{k+1}, x_{k+1} - x_k).
\end{aligned}
$$

These values only grow until the data increment $\mu$ reaches the value $\mu_0 = 2\,x_{k+1} - (x_k + x_n)$ since then $x_{k+1} - x_k \le x_n + \mu - x_{k+1}$. The double median fluctuation function is thus constant for all arguments $x > x_n + \mu_0$ when $\mu \ge \mu_0$ since it finally levels off at the MAD value, see above. This implies that every global minimizer of the double median fluctuation function lies in the interval $(x_1, x_n + \mu_0)$ even for $\mu > \mu_0$. Hence, the $MM_1$-distance equalizers remain bounded for an unbounded increase of the largest data value.

The argument also applies to growth of any other data value since that will eventually attain the role of the largest value. Yet the breakdown point generally is not larger than $1/n$ since already two data values increasing unboundedly may make the double median distance equalizer also increase unboundedly. Such examples exist for as few as five data.

#### 2.2.5 Convergence in the shape parameter

Based on the resolution of the shape parameter given in table 1, it may seem that the Lipschitz equalizers do not vary continuously in the shape parameter at the lower end. But the Lipschitz equalizers are stable there in the sense of $\lim_{p\to 1} x_{L,p} = x_{L,1}$; when any of the $L_p$-equalizers is not unique, a suitable global minimizer is chosen for the limit analysis.

# 3 Computational issues

## 3.1 Function values

A Lipschitz fluctuation function with any shape parameter can be evaluated for a given argument $x$ in $O(n)$ steps without any preprocessed results. The same applies to all modified fluctuation functions considered here.

For shape parameter one the original Lipschitz fluctuation function is piecewise linear and continuous. This allows to preprocess all points where the slope changes and to store them in a heap together with their function values. Since slope changes occur at no more than $2\,n$ points, see below, building the heap and computing all correspnding function values takes time $O(n\,\log n) + O(n^2)$.

For function evaluation at a particular argument, its right and left neighbours must be computed. The required function value is then obtained simply from linear interpolation of the two preprocessed function values. Function evaluation then requires only time $O(\log n)+O(1)$. Similar schemes apply to the median and the double median fluctuation functions since they also are piecewise linear and continuous.

## 3.2 Equalizers for two data

For the special case of only two data, their mean value is both equal to the $L_p$- and the $K_p$-equalizer for all shape parameters and equal to the $M_1$- and $MM_1$-equalizers. This is simply verified by observing that the mean value drives all fluctuation functions down to zero.

## 3.3 Equalizers with shape parameter two

Lipschitz and modified Lipschitz fluctuation functions for shape parameter two are identical quadratic functions only. Their distance equalizer can be computed for an arbitrary number of data as the unique critical point in two slightly different closed forms as

$$x_{L,2} = x_{K,2} = \frac{\sum_{i=1}^{n} x_i^2 (x_i - \bar{x})}{\sum_{i=1}^{n} 2\,x_i (x_i - \bar{x})} = \frac{\sum_{i=1}^{n} (x_i^2 - 1/n \cdot \sum_{j=1}^{n} x_j^2) \cdot (x_i - \bar{x})}{2 \sum_{i=1}^{n} (x_i - \bar{x})^2}.$$

The minimum value attained by the fluctuation function is

$$L_2(x_{L,2}) = \sum_{i=1}^{n} \Big( x_i^2 - 1/n \cdot \sum_{j=1}^{n} x_j^2 \Big)^2 + \frac{\Big( \sum_{i=1}^{n} (x_i^2 - 1/n \cdot \sum_{j=1}^{n} x_j^2) \cdot (x_i - \bar{x}) \Big)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}.$$

### 3.4 Lipschitz equalizer with shape parameter one

Lipschitz fluctuation functions for shape parameter one are piecewise linear and continuous. Minimizers hence lie at points where the function changes its slope with the only exception being an interval of constant, globally optimal function values. But then, the Lipschitz fluctuation functions change slope at both endpoints of such an interval. All slope changing points can be computed efficiently. The analysis begins with the basic expansion

$$|x-a|-|x-b| = \begin{cases} a-b, & \text{if } x < \min\{a,b\} \\ 2 \cdot sgn(b-a) \cdot x - sgn(b-a) \cdot (a+b), & \text{if } \min\{a,b\} \leq x \leq \max\{a,b\} \\ b-a, & \text{if } \max\{a,b\} < x \end{cases}$$

The slope of this function changes at the two locations $\min\{a,b\}$ and $\max\{a,b\}$ which means that changes occur at $a$ and $b$, see figure 7. Now, the absolute of the foregoing function $|\,|x-a|-|x-b|\,|$ changes slope

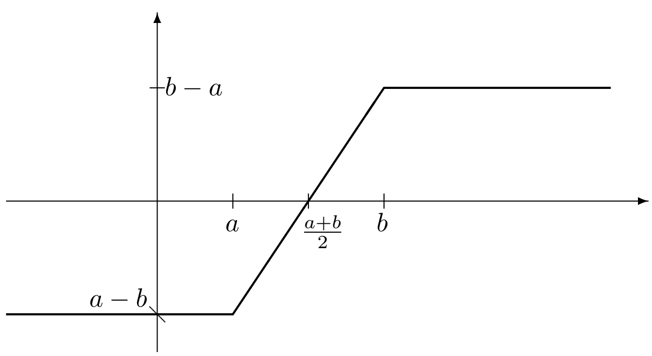


Figure 7: Function $|x-a|-|x-b|$ for $a<b$ with slope changes at $a$ and $b$ and zero at $\frac{a+b}{2}$.

at all points where the original function does and at all its zero crossings. This function has a unique zero at $\frac{a+b}{2}$ so that it changes slope at $a$, $b$, and $\frac{a+b}{2}$ with minimum value attained at $\frac{a+b}{2}$. The same analysis applies to the constituing functions

$$g_i(x) = |x_i - x| - \frac{1}{n}\sum_{j=1}^{n} |x_j - x|,$$

$i = 1, \ldots, n$. These functions generally are not monotone, change slope at the data points $x_1, \ldots, x_n$ and, tentatively, are assumed to each have a unique zero which is denoted as $z_i$. Segments on which such a function constantly attains value zero are not relevent.

Each zero crossing can be computed by sorting all data increasingly $x_1 \leq \ldots \leq x_n$ and computing all function values $g_i(x_1), \ldots, g_i(x_n)$. These values change sign at some index $k = k(i)$ which means that $sgn(g_i(x_k)) \neq sgn(g_i(x_{k+1}))$. The zero between these two data points can be computed by linear interpolation as

$$z_i = x_k + \frac{-g_i(x_k)}{g_i(x_{k+1}) - g_i(x_k)} \cdot (x_{k+1} - x_k).$$

Eventually, there remains a finite set of $n$ candidate points at which a global minimizer of the Lipschitz fluctuation function must be searched

$$
\begin{aligned}
x_{L,1} &= argmin_{x \in I\!R}\, L_1(x) = argmin_{x \in \{x_1,\ldots,x_n,z_1\ldots,z_n\}}\, L_1(x) \\
&= argmin_{x \in \{z_1\ldots,z_n\}}\, L_1(x).
\end{aligned}
$$

While it is obvious that the minimum and maximum data value can be deleted from the minimization, it is less obvious that all data can. Prooving this property is based on the observation that the sum of all constituing functions – not the sum of their absolute values – is constantly zero. Hence, for each argument, there is at least one constituing function with a positive value and another with a negative value.

The slope of the i-th constituing function decreases by $\frac{1}{n}$ at every data point with the exception of the i-th data point where "+1" is added to the slope decrement so that the slope there increases by $1 - \frac{1}{n}$. Decreases and increases refer to argument traversals from left to right. The slope increase always occurs in negative terrain since $g_i(x_i) < 0$. Beginning with slope zero to the left of the data range, the slope changes accumulate to zero over the whole data range. This agrees with the slope of every constituing function again becoming zero to the right of all data. All constituing functions have two monotonicity segments over the data range with the exception of two such functions of which one is increasing and the other is decreasing over the whole data range.

The combination of previous findings implies that the absolute of the i-th constituing function decreases its slope by $1-\frac{1}{n}$ at the i-th data point if this is not a zero of any constituing funtion. As there always is a positive and negative constituing function, the slope increment of all other absolute constituing functions is at most $(n-2)\cdot\frac{1}{n}$. Thus, the slope increment is at most $-1+\frac{1}{n}+\frac{n-2}{n} = -\frac{1}{n} < 0$ so that the fluctuation function becomes even more tilted downwards. The function can have a local minimum at $x_i$ only if its slope to the left is strictly negative or zero. But in both cases the slope to the right is strictly negative, see figure 8. Hence, the i-th data point cannot be a local minimum of the fluctuation function unless it is a zero of at least one constituing function. This completes the argument.

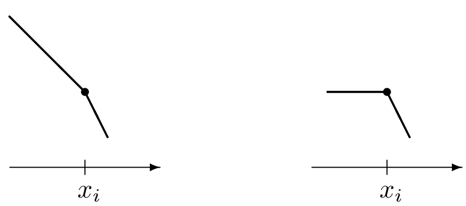


Figure 8: Fluctuation function around the i-th data point which is not a zero of any constituing function. A strictily negative slope (left) or zero slope (right) to the left of the data point is followed by a strictly negative slope to the right.

It seems that the candidate set for minimizing the fluctuation function cannot be reduced to a subset of zeros of some constituing functions only. For example, zeros that lead from positive into negative terrain and vice versa may determine the distance equalizer. A sample computation of an $L_1$-equalizer is illustrated in figure 9. While a single global minimizer is simply computed by chossing one argument with minimum objective value, computing local minima and testing for local and global minimizers forming intervals is supported by forming a sorted list of all zeros of the constituing functions.

It remains to be verified that every constituing function has only one zero. It has an odd number of zeros since the values at the maximum and minimum data have different signs; $g_i(x_n) = -g_i(x_1)$. Thus, if a constituing function had two zeros, it would have at least three zeros and, so, at least three monotonicity segments which will now be shown to be impossible. Therefore, it is noticed that the value of each constituing function at each data point as well as at the next data point are similar linear combinations of all data points

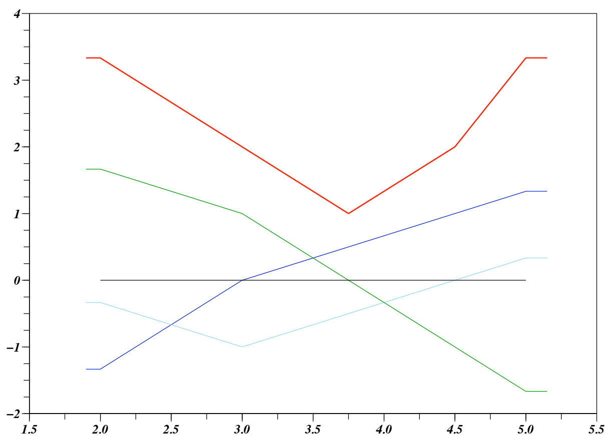


Figure 9: Graph of $L_1$-fluctuation function (upper curve) for the three data $x_1 = 2$, $x_2 = 3$, $x_3 = 5$. The zeros of the constituing functions are $z_1 = 3$ (incidently a data value), $z_2 = 4.5$, $z_3 = 3.75$ so that the distance equalizer for shape parameter one is found by enumeration over three points as $x_{L,1} = argmin_{x \in \{3, 3.75, 4.5\}} L_1(x) = 3.75$. One constituing function has two monotonicity segments while the two others each have one.

$$
\begin{array}{rcrcrcrl}
g_i(x_k) & = & A_1\,x_1 + \ldots + & A_k\,x_k & + & A_{k+1}\,x_{k+1} & + \ldots + A_n\,x_n \\
g_i(x_{k+1}) & = & A_1\,x_1 + \ldots + & (A_k - c)\,x_k & + & (A_{k+1} + c)\,x_{k+1} & + \ldots + A_n\,x_n.
\end{array}
$$

Change parameters $c$ typically are different for different values of $k$ and function values increase when the change parameter is positive. The latter can be derived as follows.

$$
\begin{aligned}
g_i(x_{k+1}) - g_i(x_k) \geq 0 \quad & \Leftrightarrow \quad (A_k - c - A_k)\,x_k + (A_{k+1} + c - A_{k+1})\,x_{k+1} \geq 0 \\
& \Leftrightarrow \quad c\,(x_{k+1} - x_k) \geq 0 \\
& \Leftrightarrow \quad c \geq 0.
\end{aligned}
$$

Monotoncity changes of a constituing function can thus be traced by sign changes in the sequence of change parameters. But each sequence of change parameters changes signs at most once as indicated in table 3.

| | $g_1$ | | sign | $g_2$ | | sign | $\ldots$ | sign | $g_n$ | sign |
|---|---|---|---|---|---|---|---|---|---|---|
| $c_1 \cdot n$ | $-2$ | $+2\,n$ | + | $-2$ | | - | | - | $-2$ | - |
| $c_2 \cdot n$ | $-4$ | $+2\,n$ | + | $-4$ | $+2\,n$ | + | | | $-4$ | - |
| $c_3 \cdot n$ | $-6$ | $+2\,n$ | + | $-6$ | $+2\,n$ | + | | | $-6$ | - |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $c_{n-1} \cdot n$ | $-2\,(n-1)$ | $+2\,n$ | + | $-2\,(n-1)$ | $+2\,n$ | + | | + | $-2\,(n-1)$ | - |

Table 3. Multiples of change parameters and their signs for each constituing function. As a reading example, a positive sign of $c_3$ indicates that $g_i(x_4)$ is larger than $g_i(x_3)$. Change parameters are multiplied by $n$ for easier notation of values. The crucial element of the table is that whenever the value $2\,n$ adds to a change parameter entry, it does so to all further change parameters of the same constituing function. Added values thus form a staircase pattern from left to right.

As a consequence of table 3 each constituing function has at most two monotonicity segments, compare figure 9, and hence a single zero between maximum and minimum data value. This is even true when some data values are equal. The arguments are summarized as the following algorithm for computing the $L_1$-distance equalizer.

$L_1$**-dist-eq**

1. Input $x_1 \leq \ldots \leq x_n$ with $x_1 < x_n$.
2. Computations
   (a) For $i = 1, \ldots, n$ do
      i. Find $k \in \{1, \ldots, n-1\}$ with $sgn(g_i(x_k)) \neq sgn(g_i(x_{k+1}))$.
      ii. $z_i = x_k + \frac{-g_i(x_k)}{g_i(x_{k+1}) - g_i(x_k)} \cdot (x_{k+1} - x_k)$.
   (b) $x_{L,1} = argmin_{x \in \{z_1 \ldots, z_n\}} L_1(x)$.
3. Output "$L_1$-distance equalizer" $x_{L,1}$.

If an index computed in step 2(a) i is not unique, then a data value is equal to a zero of the $i$-th constituing function. In this case, it suffices to select one index with the sign condition. Complete enumeration of the finite candidate set in step 2(b) requires $O(n)$ evaluations of the Lipschitz fluctuation function $L_1(x)$ which results in the overall run time complexity of $O(n^2)$.

The enumeration set of the zeros from all constituing functions can be reduced in some problem instances by a monotonicity criterion. Trivially, a zero of a constituing function is a point of equality between an absolute function $|x - x_i|$ and the mean deviation function. It suffices for some problems to consider only those equality points where an absolute function and the mean deviation function intersect at different monotonicity directions.

## 3.5 Modified Lipschitz equalizer with shape parameter one

Computations of modified Lipschitz equalizers are very similar to those of the original Lipschitz equalizers. For the present case the constituing functions become

$$g_i(x) = |x_i - x| - std(x),$$

$i = 1, \ldots, n$. These functions generally are not monotone, have curved segments between the data points and have a unique zero that may lie outside the data range. The latter is the most relevant difference to the constituing functions of ordinary Lipschitz fluctuation functions. Zeros outside the data range are irrelevant for distance equalizers. Zeros exist for all data values that are different from the mean value of the data but constituing functions for such data values need not be considered.

The arguments leading to a computation of the modified Lipschitz distance equalizer for shape parameter one are so similar to those of the ordinary Lipschitz distance equalizer that they are not repeated. In summary, they lead to the following algorithm.

$K_1$**-dist-eq**

1. Input $x_1 \leq \ldots \leq x_n$ with $x_1 < x_n$.
   Initialization $Z = \emptyset$.
2. Computations
   (a) For $i = 1, \ldots, n$

i. If $x_i \neq \bar{x}$ then $z_i = \frac{x_i^2 - 1/n \cdot \sum_{j=1}^n x_j^2}{2 \cdot (x_i - \bar{x})}$.

ii. If $x_i \neq \bar{x}$ and $z_i \in [\min\{x_1, \ldots, x_n\}, \max\{x_1, \ldots, x_n\}]$ then $Z = Z \cup \{z_i\}$.

(b) $x_{K,1} = argmin_{x \in Z} K_1(x)$.

3. Output "$K_1$-distance equalizer" $x_{K,1}$.

## 3.6 Median and double median equalizers

Computations of the single median equalizer can still be done by evaluating the zeros of constituing functions while this does no longer apply to the double median equalizer. The case of an odd number of data is concepually easier than that of an even number for both types of median equalizers and there seems to be no simple way to cover the even case by an odd number of data. In particular, inserting the median of an even set of data into that set yields an odd set but the single and the double median distance equalizers become altered by the insertion.

The constituing functions for both median cases are

$$g_i(x) = |x_i - x| - median\{|x_1 - x|, \ldots, |x_n - x|\},$$

$i = 1, \ldots, n$. These constituing functions typically do not sum to zero and zero computations of each function are different for odd and even number of data.

### 3.6.1 Single median distance equalizer

For an odd number of data, the constituing functions have segments of value zero since the median segmentwise coincides with one of the involved functions as sketched in figure 10. The endpoints of the

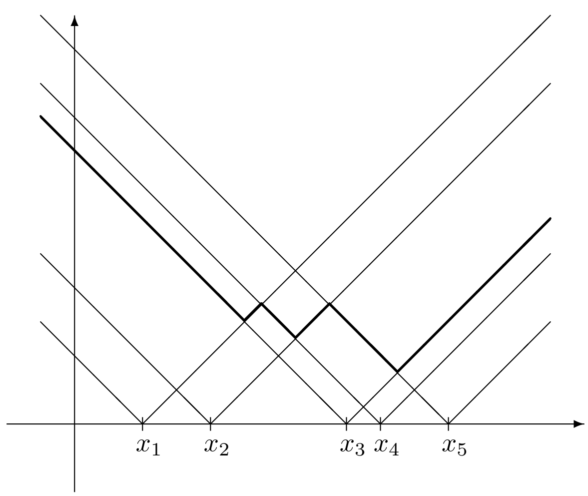


Figure 10: The median of the underlying functions $|x - x_1|, \ldots, |x - x_5|$ changes its slope where it changes its underlying function, namely at $\frac{x_1+x_3}{2}$, $\frac{x_1+x_4}{2}$, $\frac{x_2+x_4}{2}$, $\frac{x_2+x_5}{2}$ and at $\frac{x_3+x_5}{2}$.

zero segments can be computed explicitly when the data are sorted increasingly. When their number is $n = 2k - 1$, the constituing functions have the following $n$ endpoints of zero segments

$$\frac{x_i + x_{i+k-1}}{2},\ i = 1, \ldots, k \text{ and } \frac{x_i + x_{i+k}}{2},\ i = 1, \ldots, k-1.$$

A sample situation of these points is given for $n = 7$ with $k = 4$. The seven endpoints of zero segments can be sorted increasingly when upper and lower indices are increased alternatively

$$\begin{array}{cc}
\frac{x_1+x_4}{2} & \frac{x_1+x_5}{2} \\
\frac{x_2+x_5}{2} & \frac{x_2+x_6}{2} \\
\frac{x_3+x_6}{2} & \frac{x_3+x_7}{2} \\
\frac{x_4+x_7}{2} & -
\end{array}$$

When the number of data is even, the derivation of the zeros is slightly more complicated because the median is always formed as mean value of two middle values. As a consequence, the constituing functions typically have no zero segments as shown in figure 11 and each zero is attained by two constituing functions. Generally, the zeros for an even number of data $n = 2\,k$ are the $n/2$ values

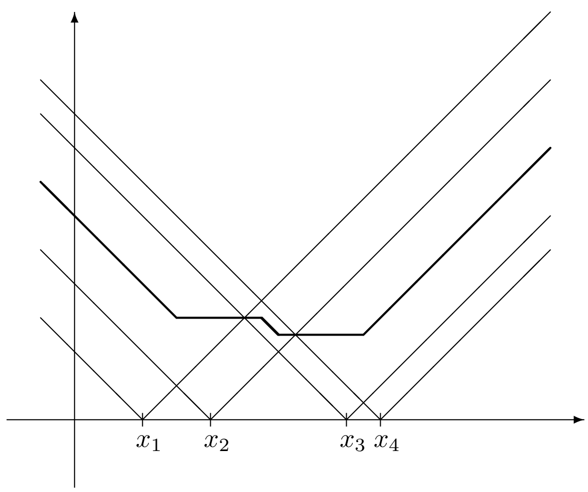


Figure 11: The median of the underlying functions $|x - x_1|, \ldots, |x - x_4|$ changes its slope where it changes its underlying function. Zeros of the constituing functions lie at $\frac{x_1+x_3}{2}$ and $\frac{x_2+x_4}{2}$.

$$\frac{x_1 + x_{1+k}}{2}, \frac{x_2 + x_{2+k}}{2}, \ldots, \frac{x_k + x_n}{2}.$$

None of these zeros is of the general form $median\{x_1, \ldots, x_{2k}\} = \frac{x_k + x_{1+k}}{2}$, so that the single distance equalizer will always be different from the median of an even number of pairwise different data. The overall procedure for the single median equalizer can thus be summarized as follows.

$M_1$**-dist-eq**

1. Input $x_1 \le \ldots \le x_n$ with $x_1 < x_n$.
2. Computations
   (a) If $n$ is odd with $n = 2\,k - 1$ then $Z = \{\frac{x_i + x_{i+k-1}}{2},\ i = 1, \ldots, k \text{ and } \frac{x_i + x_{i+k}}{2},\ i = 1, \ldots, k-1\}$.
   If $n$ is even with $n = 2\,k$ then $Z = \{\frac{x_i + x_{i+k}}{2},\ i = 1, \ldots, k\}$.

(b) $x_{M,1} = argmin_{x \in Z}\, M_1(x)$.

3. Output "$M_1$-distance equalizer" $x_{M,1}$.

### 3.6.2 Double median distance equalizer

As a break from algorithms for all foregoing distance equalizers, double median equalizers cannot be computed from the zeros of their constituing functions as shown in figure 12. This is a consequence of the double median distance equalizer being the only equalizer whose fluctuation function is not aggregated by summation. A closer analysis reveals that the intersection points of distinct absolute constituing functions must be considered in addition to their zeros or endpoints of their zero segments. As endpoints of zero segments are special cases of intersection points of distinct absolute constituing functions, the set of candidate points is only generalized mildly from the current conceptual viewpoint.

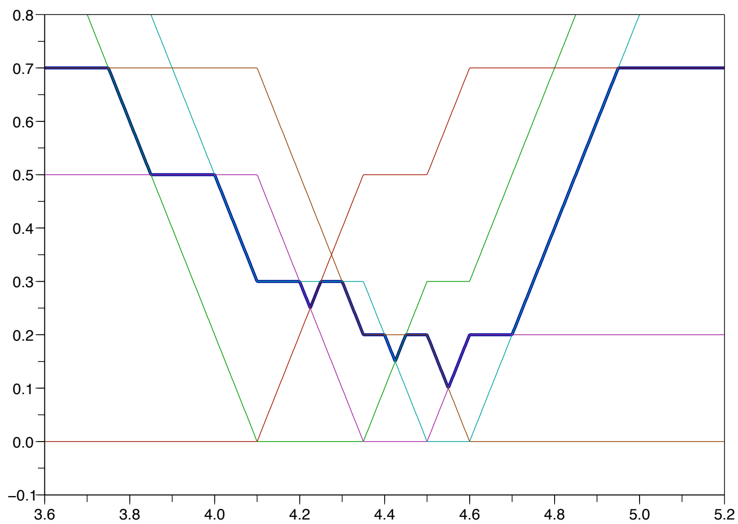


Figure 12: Partial graphs of the $MM_1$-fluctuation function (bold curve) with global minimum at 4.55 and partial graphs of the absolute constituing functions for the five data $2.7, 3, 5.5, 6, 6.2$. None of the absolute constituing functions has an endpoint of a zero segment at the global or at the two local minima but function intersections occur there.

The formal intersection condition to be considered is $|g_a(x)| = |g_b(x)|$ for distinct indices $a, b$. Omitting the absolute values gives either $g_a(x) = g_b(x)$ which simply implies $x = \frac{x_a + x_b}{2}$ or it gives $g_a(x) = -g_b(x)$ which implies

$$\frac{|x_a - x| + |x_b - x|}{2} = median\{|x_1 - x|, \ldots, |x_n - x|\}.$$

A difficulty associated with solutions of this equation is that the solution set may be disconnected. This warrants a search procedure as opposed to closed form solutions. Assuming that the number of data is odd results in the median being attained by a single underlying function so that

$$\frac{|x_a - x| + |x_b - x|}{2} = |x_i - x|.$$

Solutions of this equation can be computed by first generating candidate values and then verifying if these actually are medians. Solution candidates depend on the three indices as follows

$$x_{i,a,b} = \begin{cases} x_i - \frac{|x_a - x_b|}{2}, & \text{if } \ x_i \in [\frac{x_a+x_b}{2}, A_{a,b}) \\ x_i + \frac{|x_a - x_b|}{2}, & \text{if } \ x_i \in (B_{a,b}, \frac{x_a+x_b}{2}] \\ \max\{x_a, x_b\} + \frac{x_i - A_{a,b}}{2}, & \text{if } \ x_i > A_{a,b} \\ \min\{x_a, x_b\} - \frac{B_{a,b} - x_i}{2}, & \text{if } \ x_i < B_{a,b} \end{cases}$$

with interval bounds

$$A_{a,b} = \frac{x_a + x_b}{2} + |x_a - x_b| \text{ and } B_{a,b} = \frac{x_a + x_b}{2} - |x_a - x_b|.$$

If such a candidate solution satisfies the median condition $median\{|x_{i,a,b} - x_1|, \ldots, |x_{i,a,b} - x_n|\} = |x_{i,a,b} - x_i|$, then it is retained for further inspection, otherwise it is discarded. Index combinations may be redundant but some intersections are reached by a unique combination and all three indices of such a combination may be different.

Assuming now that the number of data is even results in the median being attained by the mean value of two underlying functions so that the intersection condition for constituing functions $g_a(x) = -g_b(x)$ becomes

$$\frac{|x_a - x| + |x_b - x|}{2} = \frac{|x_i - x| + |x_j - x|}{2}$$

for $a \neq b$ and $i \neq j$. Similar formulas than the preceeding lead to candidate solutions $x_{i,j,a,b}$ for all different indices. Once of of the first two indices is equal to one of the last two indices, their contributions from the equation cancel out and the average of the data remains. Candidate solutions must satisfy the median condition in order to be retained for optimization. All together, these operations result in an algorithm for double median distance equalizers.

$MM_1$**-dist-eq**

1. Input $x_1 \leq \ldots \leq x_n$ with $x_1 < x_n$.
   Initialization $Z = \{\frac{x_i+x_j}{2}, \ 1 \leq i < j \leq n\}$.
2. Computations
   (a) For $i = 1, \ldots, n$
   If $n$ is odd and $\Big(z$ is an endpoint of a zero segment of function $g_i(x)$ or $z = x_{i,a,b} \in (x_1, x_n)$ with $a, b \in \{1, \ldots, n\} - \{i\}$ satisfies $median(|x_{i,a,b} - x_1|, \ldots, |x_{i,a,b} - x_n|) = |x_{i,a,b} - x_i|\Big)$ then $Z = Z \cup \{z\}$.
   If $n$ is even and $\Big(z$ is a zero of function $g_i(x)$ or $z = x_{i,j,a,b} \in (x_1, x_n)$ with $a, b \in \{1, \ldots, n\} - \{i, j\}$ satisfies $median(|x_{i,j,a,b} - x_1|, \ldots, |x_{i,j,a,b} - x_n|) = (|x_{i,j,a,b} - x_i| + |x_{i,j,a,b} - x_j|)/2\Big)$ then $Z = Z \cup \{z\}$.
   (b) $x_{MM,1} = argmin_{x \in Z} \, MM_1(x)$.
3. Output "$MM_1$-distance equalizer" $x_{MM,1}$.

### 3.7 Finite shape parameters different from one and two

A reasonable strategy to compute equalizers for a shape parameter different from one and two is to use computationally tractable distance equalizers as initial values for approximation algorithms. Therefore, the equalizers for shape parameter two is computed as well as all equalizers for shape parameter one. The initial equalizers are complemented by the minimum and the maximum value of the data. Any adjacent pair of these values is then selected for bisection over the subinterval formed by the two values. During all bisections, the fluctuation function with shape parameter in question is used. At least three candidate locations result from the bisections and their best is retained as approximation of the equalizer. Both $L_p$- and $K_p$-equalizers can be dealt with in this way.

### 3.8 Shape parameter infinity

The distance equalizer for shape parameter infinity can simply be computed as the midpoint of the data range

$$x_{L,\infty} = \frac{\max\{x_1, \ldots, x_n\} + \min\{x_1, \ldots, x_n\}}{2}.$$

Though alternative global minimizers of the fluctuation function $L_\infty(x)$ thus cannot captured, the simplicity of the computation warrants its application. The midpoint of the data range may lie in the interior or at one of the endpoints of the interval of global minimizers. A derivation of the equalizer formula can be based on the equations

$$\begin{aligned}
argmin_{x\in I\!R}\, L_\infty(x) &= argmin_{x\in I\!R} \max_{i=1,\ldots,n} \left| \max_{j=1,\ldots,n} |x_j - x| - |x_i - x| \right| \\
&= argmin_{x\in I\!R} \max_{i=1,\ldots,n} \left( \max_{j=1,\ldots,n} |x_j - x| - |x_i - x| \right) \\
&= argmin_{x\in I\!R} \left( \max_{j=1,\ldots,n} |x_j - x| - \min_{i=1,\ldots,n} |x_i - x| \right) \\
&= \frac{x_1 + x_n}{2}.
\end{aligned}$$

for sorted data $x_1 \leq \ldots \leq x_n$. The outer maximization is attained as difference between maximum and minimum. Validity of the last equality is illustrated by figure 13. It can also be deduced from this figure that the $L_\infty$-equalizer is unique if the minimum of the maximum function coincides with a local maximum of the minimum function. The condition is inferred from leaving the maximum and minimum data fixed and slightly shifting all other data to the right.

## 4 Fluctuation clustering

Clustering data by fluctuation is undersood, in complete analogy to other clustering approaches, as finding a partition of a given data set with minimum sum of fluctuation values for a given number of classes. Classes $X_1, \ldots, X_m$ of one clustering are pairwise disjoint and cover the data set; $X_i \cap X_j = \emptyset$ for $i \neq j$ and $X_1 \cup \ldots \cup X_m = \{x_1, \ldots, x_n\}$. Finding an optimal clustering with $m$ classes can refer to all types of fluctuation functions considered so far but, in order to avoid too many repetitions, only ordinary Lipschitz fluctuation functions with finite shape parameters will now be dealt with. This amounts to solving the minimization problem

$$\min_{X_1,\ldots,X_m} \sum_{i=1}^{m} L_p(X_i).$$

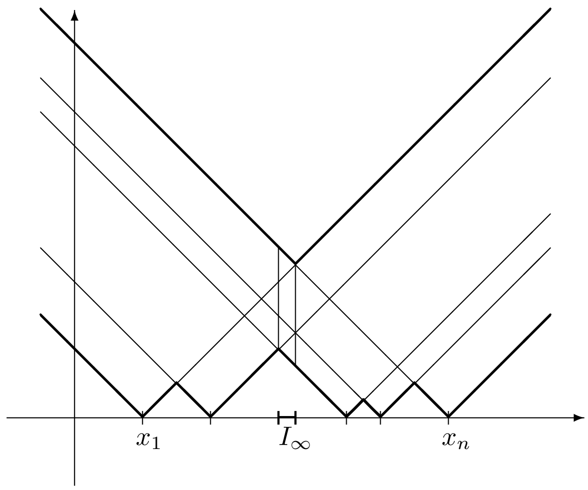


Figure 13: The difference between the maximum and the minimum of the functions $|x - x_1|, \ldots, |x - x_n|$ is minimized between the two vertical line segments which have equal length. Here, the right endpoint of the minimum interval $I_\infty$ is the midpoint of the data range.

The fluctuation value of a class is defined as the minimum value of the fluctuation function attainable there. Obviously, this is the function value of the distance equalizer or dispersion value over that class. For a class $X \subseteq \{x_1, \ldots, x_n\}$ its fluctuation value is

$$L_p(X) = \min_{x \in \mathbb{R}} \sum_{x_k \in X} \left| |x - x_k|^p - \frac{1}{|X|} \sum_{x_j \in X} |x - x_j|^p \right|^p.$$

Trivially, the fluctuation value of a singleton is zero but it is also zero for any class of two points, see section 3.2. Hence, it is assumed that the class number is reasonably low like $m < n/2$, since clusterings with more classes can be formed of singletons and pairs only resulting in fluctuation value zero of the whole clustering. Yet clusterings that are considered along computations may have more classes.

As a preparatory step towards optimal clustering, a best triplet class is constructed. This proceeds in the same way for all shape parameters by, first, finding two data with minimum distance between them and, second, finding a third data point with minimum distance to the center of the two. Ties in the first step must not be broken arbitrarily but each pair of minimum distance must be inspected together with any third point coming closest to the pair center. Only ties remaining after the selection of the third point may be broken arbitrarily.

An optimal fluctuation clustering may have classes that comprise other classes. For example, an optimal clustering of the six data $0, 10, 15, 16, 16.1, 40$ with two classes and shape parameter one has the partition $X_1 = \{10, 15, 16, 16.1\}$, $X_2 = \{0, 40\}$. The convex hull of one class is strictly contained in the convex hull of the other class. In principle, such inclusions can be avoided by the transition to an alternative, optimal clustering. In the example, the classes $X_1 = \{0, 10\}$, $X_2 = \{15, 16, 16.1, 40\}$ are also optimal and have disjoint convex hulls.

Anyway, if the shape parameter is different from one, the transition to classes whose convex hulls are disjoint may lead away from optimality. The clustering $X_1 = \{10, 15, 16, 16.1\}$, $X_2 = \{0, 40\}$ is uniquely optimal for shape parameter two so that, in particular, it is optimal to group the two data with maximum distance in the same class.

# 5 Cluster computations

## 5.1 Heuristic computations

Computations of good fluctuation clusterings may either have to backtrack, which means revising previous decisions, or have to wrap around. The latter means that they may have to arrange large and small data in one class as in the foregoing example. Heuristic computations, which should be simple and, so, avoid backtracking, must be ready to wrap around. To facilitate this, all data are symbolically arranged on a cycle which means that the largest data value is not only a neighbour of the second largest but also to the smallest.

A heuristic may then proceed from an all singleton clustering by successively selecting one of the following operations until the desired class number is reached:

- unite two cyclic-neighbouring classes of which at least one contains at least three data,
- unite three cyclic-neighbouring singletons to one class,
- if only two cyclic-neighbouring singletons are left, unite them.

The operations are selected in a greedy manner: whenever operations are applicable to several choices of classes, one choice leading to the least sum of fluctuation values is selected. Possible operations on the cyclic data arrangement are sketched in figure 14.

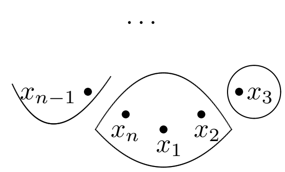


Figure 14: Sketch of one class of three data which are wrapped around the data range, one singleton and another class containing $x_{n-1}$.

## 5.2 Cluster graphs

Computing fluctuation clusterings corresponds to computing shortest paths with prescribed number of intermediate nodes in certain weighted acyclic graphs. These graphs are here defined as cluster graphs. Their nodes are denoted as support points which indicate where two classes meet. Thus, support points interleave the data points and one support point is added to the left and another to the right of all data points. The arrangement of all support points and all data points is $z_0 < x_1 < z_1 < x_2 < z_2 < \ldots < x_n < z_n$ as indicated in figure 15. Noteably, the particular position of a support point does not matter.


Figure 15: Support points interleaved with the data points.

Arcs are now introduced to lead from any support point to every other with greater index; arcs are ordered pairs of nodes $(z_i, z_j)$ for all $0 \leq i < j \leq n$. The resulting cluster graph is a complete directed acyclic graph. An arc $(z_i, z_j)$ denotes one class $X$ of a clustering, namely that of all data points $x_{i+1}, \ldots, x_j$ which lie between $z_i$ and $z_j$ as indicated in figure 16. Arcs are assigned weights $w(z_i, z_j) = L_p(\{x_{i+1}, \ldots, x_j\})$. All weights are tentatively assumed to be known and they likely violate the triangle inequality.

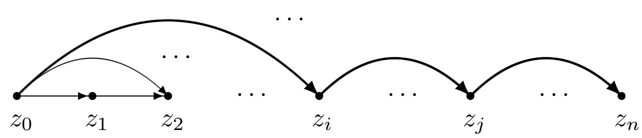


Figure 16: Sketch of a cluster graph with indication of a path for a classification into three claases separated at $z_i$ and $z_j$.

Now, the pivotal observation is as follows. A clustering of $m$ classes without wrap around corresponds to a path from node $z_0$ to node $z_n$ with exactly $m-1$ intermediate nodes. The objective value of a particular classification is the cost of its path. Thus, an optimal classification with $m$ classes without wrap around corresponds to a shortest path from $z_0$ to $z_n$ in the cluster graph with exactly $m-1$ intermediate nodes. Such a path can be found by dynamic programming as well as by a modification of the Dijkstra algorithm which is given next.

**Mod-Dijk**

1. Input.
   - (a) Cluster graph with nodes $z_0, \ldots, z_n$.
   - (b) Weights $w(z_i, z_j)$ for $0 \leq i < j \leq n$.
   - (c) Desired class number $m \geq 2$.

   Initialization.
   - (a) $D_i(k) = w(z_0, z_i)$ for $i = 1, \ldots, n$ and $k = 0, \ldots, m-1$,
   - (b) $T_k = \{1, \ldots, n\}$ for $k = 1, \ldots, m-1$.
2. Iteration. For $k = 1, \ldots, m-1$ do

   While candidate list $T_k \neq \emptyset$ do
   - (a) Compute $i = argmin_{j \in T_k} D_j(k)$.
   - (b) $T_k = T_k - \{i\}$.
   - (c) $\forall\, j \in T_k$ compute $pred(j) = i$, if $D_i(k-1) + w(z_i, z_j) < D_j(k)$.
   - (d) $\forall\, j \in T_k$ compute $D_j(k) = min\{D_i(k-1) + w(z_i, z_j), D_j(k)\}$.
3. Output.
   - (a) Optimal fluctuation clustering value $D_n(m-1)$.
   - (b) Decreasingly sorted support point indices $pred(n), pred(pred(n)), \ldots, pred^{(m-1)}(n) = w_{m-1}, \ldots, w_1$.

The decreasing index sequence generated by algorithm **Mod-Dijk** results in the classes $X_1 = \{x_1, \ldots, x_{w_1}\}$, $X_2 = \{x_{w_1+1}, \ldots, x_{w_2}\}, \ldots, X_m = \{x_{w_{m-1}+1}, \ldots, x_n\}$. The algorithm has run time complexity $O(m\, n^2)$.

Admitting wrap around makes cluster computations more complicated since it must be determined which data belong to the wrap around class for both the lower and the upper area of the data range. This can be done by modifying the cluster graph and the shortest path computations.

Altering the path computations will account for the lower end of the data range. When assigning the smallest data point to the wrap around class, path computations will begin at $z_1$ instead of $z_0$. When assigning the two smallest data points to the wrap around class, path computations will begin at $z_2$ instead of $z_0$ etc. Different cases of such jumps have to be considered.

Altering arc labels will account for the upper end of the data range. All arcs that lead to the terminal node will be relabeled so that their weights include all data points which were jumped over. This means that the arcs $(z_i, z_n)$, $i = 1, \ldots, n-1$, are relabled from $L_p(\{x_{i+1}, \ldots, x_n\}$ to $L_p(\{x_1, \ldots, x_k, x_{i+1}, \ldots, x_n\}$ if the first $k$ data points are jumped over. Cluster compuations are then enabled by the following shortest path computations that use algorithm **Mod-Dijk** in each iteration with modified input.

**Cluster**

1. Input.
   (a) Cluster graph with nodes $z_0, \ldots, z_n$.
   (b) Weights $w(z_i, z_j)$ for $0 \leq i < j \leq n$.
   (c) Desired class number $m \geq 2$ and maximum number of data jumped over at lower end of data range $0 \leq K < n$.

   Initialization.

   $record\_length = \infty$.

2. Iteration. For $k = 0, \ldots, K$ do
   (a) Compute shortest path $P_k$ from $z_k$ to $z_n$ with $m-1$ intermediate nodes by **Mod-Dijk**.
   (b) If $length(P_k) < record\_length$ then $P = P_k$ and $record\_length = length(P_k)$.
   (c) Assign label $L_p(\{x_1, \ldots, x_k, x_{i+1}, \ldots, x_n\})$ to arc $(z_i, z_n)$, $i = 1, \ldots, n-1$.

3. Output. Shortest of all shortest paths $P$.

Classes of the optimal clustering can be computed from the output path similar to the class computations from the output path of the modified Dijkstra algorithm. It remains the difficulty to compute the arc weights of the cluster graph. This will be sketched for Lipschitz fluctuation functions with shape parameters one and two only.

## 5.3 Arc weights for shape parameter two

Computing all arc weights required by algorithm **Cluster** can be achieved for shape parameter two by the closed formula for minimum values of the fluctation function, see section 3.3. The computations of various classes can then be interleaved so that computing all $O(n^2)$ arc weights of the cluster graph requires computing effort $O(n^2)$ and the update effort in each iteration of step 2(c) of algorithm **Cluster** requires effort $O(n)$. Interleaved computations are enabled by the expansion

$$L_2(X) = D - \frac{B^2}{|X|} + \frac{(C - AB/|X|)^2}{B - A^2/|X|}$$

with sums of the first four powers over all data from any class $X$

$$A = \sum_{x_i \in X} x_i,\ B = \sum_{x_i \in X} x_i^2,\ C = \sum_{x_i \in X} x_i^3,\ D = \sum_{x_i \in X} x_i^4.$$

### 5.4 Arc weights for shape parameter one

Computing the arc weights for the cluster graph so that intermediate results can be reused does not seem possible for shape parameter one. This results in quite a heavy preprocessing load to enable proper clustering.

## References


[Fa] Fabian, Z., "New measures of central tendency and variability of continuous distributions", Communications in Statistics - Theory and Methods 37, 2008, p. 159-174.

[Ho] Hosking, J.R.M., "L-moments: analysis and estimation of distributions using linear combinations of order statistics", Journal of the Royal Statistical Society B, 52, 1990, p. 105-124.

[K] Kämpke, T., "Signal quantization by variation", submitted, 2010.

[LoRo] Lopuhaä, H.P., Rousseeuw, P.J., "Breakdown points of affine equivariant estimators of multivariate location and covariance matrices", The Annals of Statistics 19, 1991, p. 229-248.

[MaBr] Maindonald, J., Braun, J., "Data analysis and graphics using R: an example-based approach", Cambridge University Press, Cambridge, 2006.

[MOPP] Marchand., E., Ouassou, I., Payandeh, A.T., Perron, F., "On the estimation of a restricted location parameter for symmetric distributions", Journal of the Jap. Statistical Society 38, 2008, p. 293-309.

[Sci] Scilab, Computer algebra system, www.scilab.org.

[Sh] Shakarji, C.M., "Least-squares fitting algorithms of the NIST algorithm testing system", Journal of Research of the National Institute of Standards and Technology 103, 1998, p. 633-641.

[Si] Silver, G.L., "Numerical relations within cubes and the estimation of central tendency", Applied Mathematical Sciences 3, 2009, p. 119-128.