Fitting a Graph to One-Dimensional Data

Siu-Wing Cheng\textsuperscript{1}  Otfried Cheong\textsuperscript{2}  Taeyoung Lee\textsuperscript{2}

September 11, 2018

Abstract

Given \( n \) data points in \( \mathbb{R}^d \), an appropriate edge-weighted graph connecting the data points finds application in solving clustering, classification, and regression problems. The graph proposed by Daitch, Kelner and Spielman (ICML 2009) can be computed by quadratic programming and hence in polynomial time. While in practice a more efficient algorithm would be preferable, replacing quadratic programming is challenging even for the special case of points in one dimension. We develop a dynamic programming algorithm for this case that runs in \( O(n^2) \) time. Its practical efficiency is also confirmed in our experimental results.

1 Introduction

Many interesting data sets can be interpreted as point sets in \( \mathbb{R}^d \), where the dimension \( d \) is the number of features of interest of each data point, and the coordinates are the values of each feature. To model the similarity between discrete samples, one can introduce appropriate undirected weighted edges connecting proximal points. Such a graph is useful in applications such as classification, regression, and clustering (see, for instance, \cite{5, 8}). For example, let \( w_{ij} \) denote the weight determined for the edge that connects two points \( p_i \) and \( p_j \), and regression can be performed to predict function values \( f_i \)'s at the points \( p_i \)'s by minimizing \( \sum_{i,j} w_{ij} (f_i - f_j)^2 \), subject to fixing the subset of known \( f_i \)'s \cite{1}. As another example, for any given integer \( k \), one can obtain a partition of the weighted graph into \( k \) clusters based on spectral analysis of the eigenvectors of the Laplacian of the weighted graph \cite{1, 5}. Note that the weighted graph may actually be connected. To allow efficient data analysis, it is important that the weighted graph is sparse.

Different proximity graphs have been suggested for this purpose. The \( k\text{NN} \)-graph connects each point to its \( k \) nearest neighbors. The \( \varepsilon \)-ball graph connects each point to all other points that are within a distance \( \varepsilon \). In both cases, an edge of length \( \ell \) is assigned a weight of \( \exp(-\ell^2/2\sigma^2) \), where the parameters \( k, \varepsilon \) and \( \sigma \) need to be specified by the user. It is unclear how to set these parameters in an automatic, efficient way. Several studies have found the \( k\text{NN} \)-graph and the \( \varepsilon \)-ball graphs to be inferior to other graphs proposed \cite{1, 2, 7}.

We consider the graph proposed by Daitch, Kelner, and Spielman \cite{1}. It is provably sparse, and experiments have shown that it offers good performance in classification, clustering and regression. This graph is defined via quadratic optimization as follows: Let \( P = \{p_1, p_2, \ldots, p_n\} \) be a set of

\textsuperscript{1}Supported by Research Grants Council, Hong Kong, China (project no. 16200317). Department of Computer Science and Engineering, HKUST, Clear Water Bay, Hong Kong. Email: scheng@cse.ust.hk

\textsuperscript{2}Supported by ICT R&D program of MSIP/ITTP [R0126-15-1108]. School of Computing, KAIST, Daejeon, South Korea. Email: otfried@kaist.airpost.net, taeyoung@kaist.ac.kr
\( n \) points in \( \mathbb{R}^d \). We assign weights \( w_{ij} \geq 0 \) to each pair of points \((p_i, p_j)\), such that \( w_{ij} = w_{ji} \) and \( w_{ii} = 0 \). These weights determine for each point \( p_i \) a vector \( \vec{v}_i \), as follows:

\[
\vec{v}_i = \sum_{j=1}^{n} w_{ij} (p_j - p_i).
\]

Let \( v_i \) denote \( \| \vec{v}_i \| \). The weights are chosen so as to minimize the sum

\[
Q = \sum_{i=1}^{n} v_i^2,
\]

under the constraint that the weights for each point add up to at least one (to prevent the trivial solution of \( w_{ij} = 0 \) for all \( i \) and \( j \)):

\[
\sum_{j=1}^{n} w_{ij} \geq 1 \quad \text{for } 1 \leq i \leq n.
\]

The resulting graph contains an edge connecting \( p_i \) and \( p_j \) if and only if \( w_{ij} > 0 \).

Daitch et al. [1] showed that there is an optimal solution where at most \((d + 1)n\) weights are non-zero. Moreover, in two dimensions, optimal weights can be chosen such that the graph is planar.

Clearly, the optimal weights can be computed by quadratic programming. A quadratic programming problem with \( m \) variables, \( c \) constraints, and \( L \) input bits can be solved in \( O(m^4L^2) \) time using the method of Ye and Tse [6]. There is another algorithm by Kapoor and Vaidya [4] that has an asymptotic running time of \( O((m + c)^{3.67}L \cdot \log L \cdot \log(m + c)) \). In our case, there are \( n(n - 1)/2 \) variables and \( \Theta(n) \) constraints. So the running time is \( O(n^{7.34}L \cdot \log L \cdot \log n) \), which is impractical even for moderately large \( n \). Daitch et al. reported that a data set of 4177 points requires a processing time of approximately 13.8 hours. Graphs based on optimizing other convex quality measures have also been considered [3, 7].

Our goal is to design an algorithm to compute the optimal weights in Daitch et al.’s formulation that is significantly faster than quadratic programming. Perhaps surprisingly, this problem is challenging even for points in one dimension, that is, when all points lie on a line. In this case, it is not difficult to show (Lemma 2.1) that there is an optimal solution such that \( w_{ij} > 0 \) if and only if \( p_i \) and \( p_j \) are consecutive. This reduces the number of variables to \( n - 1 \). Even in one dimension, the weights in an optimal solution do not seem to follow any simple pattern as we illustrate in the following two examples.

Some weights in an optimal solution can be arbitrarily high. Consider four points \( p_1, p_2, p_3, p_4 \) in left-to-right order such that \( \|p_1 - p_2\| = \|p_3 - p_4\| = 1 \) and \( \|p_2 - p_3\| = \varepsilon \). By symmetry, \( w_{12} = w_{34} \), and so \( v_1 = v_4 = w_{12} \). Since \( w_{12} + w_{23} \geq 1 \) and \( w_{23} + w_{34} \geq 1 \) are trivially satisfied by the requirement that \( w_{12} = w_{34} \geq 1 \), we can make \( v_2 \) zero by setting \( w_{23} = w_{12}/\varepsilon \). In the optimal solution, \( w_{12} = w_{34} = 1 \) and \( w_{23} = 1/\varepsilon \). So \( w_{23} \) can be arbitrarily large.

Given points \( p_1, \cdots, p_n \) in left-to-right order, it seems ideal to make \( v_i \) a zero vector. One can do this for \( i \in [2, n - 1] \) by setting \( w_{i-1,i}/w_{i,i+1} = \|p_i - p_{i+1}\|/\|p_{i-1} - p_i\| \), however, some of the constraints \( w_i + w_{i+1} \geq 1 \) may be violated. Even if we are lucky that for \( i \in [2, n - 1] \), we can set \( w_{i-1,i}/w_{i,i+1} = \|p_i - p_{i+1}\|/\|p_{i-1} - p_i\| \) without violating \( w_i + w_{i+1} \geq 1 \), the solution may not be optimal as we show below. Requiring \( v_i = 0 \) for \( i \in [2, n - 1] \) gives \( v_1 = v_n = w_{12}\|p_1 - p_2\| \). In
general, we have \(\|p_1 - p_2\| \neq \|p_{n-1} - p_n\|\), so we can assume that \(\|p_1 - p_2\| > \|p_{n-1} - p_n\|\). Then, \(w_{n-1,n} = w_{12}\|p_1 - p_2\|/\|p_{n-1} - p_n\| > 1\) as \(w_{12} \geq 1\). Since \(w_{n-1,n} > 1\), one can decrease \(w_{n-1,n}\) by a small quantity \(\delta\) while keeping its value greater than 1. Both constraints \(w_{n-1,n} \geq 1\) and \(w_{n-2,n-1} + w_{n-3,n-2} + w_{n-2,n-1} > 1\) are still satisfied. Observe that \(v_n\) drops to \(w_{12}\|p_1 - p_2\| - \delta\|p_{n-1} - p_n\|\) and \(v_{n-1}\) increases to \(\delta\|p_{n-1} - p_n\|\). Hence, \(v_{n-1}^2 + v_n^2\) decreases by \(2\delta w_{12}\|p_1 - p_2\|\|p_{n-1} - p_n\| - 2\delta^2\|p_{n-1} - p_n\|^2\), and so does \(Q\). The original setting of the weights is thus not optimal. If \(w_{n-3,n-2} + w_{n-2,n-1} > 1\), it will bring further benefit to decrease \(w_{n-2,n-1}\) slightly so that \(v_{n-1}\) decreases slightly from \(\delta\|p_{n-1} - p_n\|\) and \(v_{n-2}\) increases slightly from zero. Intuitively, instead of concentrating \(w_{12}\|p_1 - p_2\|\) at \(v_n\), it is better to distribute it over multiple points in order to decrease the sum of squares. But it does not seem easy to determine the best weights.

Although there are only \(n - 1\) variables in one dimension, quadratic programming still yields a high running time of \(O(n^{3.67}L \cdot \log L \cdot \log n)\). We present a dynamic programming algorithm that computes the optimal weights in \(O(n^2)\) time in the one-dimensional case. The intermediate solution has an interesting structure such that the derivat ure of its quality measure depends on the derivative of a subproblem’s quality measure as well as the inverse of this derivative function. This makes it unclear how to bound the size of an explicit representation of the intermediate solution. Instead, we develop an implicit representation that facilitates the dynamic programming algorithm.

We implemented our algorithm, with both the explicit and the implicit representation of intermediate solutions. Both versions run substantially faster than the quadratic solver in CVXOPT. For instance, for 3200 points, CVXOPT needs over 20 minutes to solve the quadratic program, while our algorithm takes less than half a second to compute the optimal weights.

## 2 A single-parameter quality measure function

We will assume that the points are given in sorted order, so that \(p_1 < p_2 < p_3 < \cdots < p_n\). We first argue that the only weights that need to be non-zero are the weights between consecutive points, that is, weights of the form \(w_{i,i+1}\).

**Lemma 2.1.** For \(d = 1\), there is an optimal solution where only weights between consecutive points are non-zero.

**Proof.** Assume an optimal solution where \(w_{ik} > 0\) and \(i < j < k\). We construct a new optimal solution as follows: Let \(a = p_j - p_i\), \(b = p_k - p_j\), and \(w = w_{ik}\). In the new solution, we set \(w_{ik} = 0\), increase \(w_{ij}\) by \(\frac{a+b}{a}w\), and increase \(w_{jk}\) by \(\frac{a+b}{b}w\). Note that since \(a + b > a\) and \(a + b > b\), the sum of weights at each vertex increases, and so the weight vector remains feasible. The value \(v_j\) changes by \(-a \times \frac{a+b}{a}w + b \times \frac{a+b}{b}w = 0\), the value \(v_i\) changes by \(-(a+b) \times w + a \times \frac{a+b}{a}w = 0\), and the value \(v_k\) changes by \((a+b) \times w - b \times \frac{a+b}{b}w = 0\). It follows that the new solution has the same quality as the original one, and is therefore also optimal. \(\square\)

To simplify the notation, we set \(d_i = p_{i+1} - p_i\), for \(1 \leq i < n\), rename the weights as \(w_i := w_{i,i+1}\), again for \(1 \leq i < n\), and observe that

\[
\begin{align*}
v_1 &= w_1 d_1, \\
v_i &= |w_i d_i - w_{i-1} d_{i-1}| \quad \text{for } 2 \leq i \leq n-1, \\
v_n &= w_{n-1} d_{n-1}.
\end{align*}
\]
For $i \in [2, n-1]$, we introduce the quantity
\[ Q_i = d_i^2 w_i^2 + \sum_{j=1}^{i} v_j^2 = d_i^2 w_i^2 + d_i^2 w_1^2 + \sum_{j=2}^{i} (d_j w_j - d_{j-1} w_{j-1})^2, \]
and note that $Q_{n-1} = \sum_{i=1}^{n} v_i^2 = Q$. Thus, our goal is to choose the $n-1$ non-negative weights $w_1, \ldots, w_{n-1}$ such that $Q_{n-1}$ is minimized, under the constraints
\[
\begin{align*}
& w_1 \geq 1, \\
& w_j + w_{j+1} \geq 1 \quad \text{for } 2 \leq j \leq n-2, \\
& w_{n-1} \geq 1.
\end{align*}
\]

The quantity $Q_i$ depends on the weights $w_1, w_2, \ldots, w_i$. We concentrate on the last one of these weights, and consider the function
\[ w_i \mapsto Q_i(w_i) = \min_{w_1, \ldots, w_{i-1}} Q_i, \]
where the minimum is taken over all choices of $w_1, \ldots, w_{i-1}$ that respect the constraints $w_1 \geq 1$ and $w_j + w_{j+1} \geq 1$ for $2 \leq j \leq i-1$. The function $Q_i(w_i)$ is defined on $[0, \infty)$.

We denote the derivative of the function $w_i \mapsto Q_i(w_i)$ by $R_i$. We will see shortly that $R_i$ is a continuous, piecewise linear function. Since $R_i$ is not differentiable everywhere, we define $S_i(x)$ to be the right derivative of $R_i$, that is
\[ S_i(x) = \lim_{y \to x^+} R_i'(y). \]

The following theorem discusses $R_i$ and $S_i$. The shorthand $\xi_i := 2d_i d_{i+1}$, for $1 \leq i \leq n-1$, will be convenient in its proof and the rest of the paper.

**Theorem 2.1.** The function $R_i$ is strictly increasing, continuous, and piecewise linear on the range $[0, \infty)$. We have $R_i(0) < 0$, $S_i(x) \geq (2 + \frac{2}{i})d_i^2$ for all $x \geq 0$, and $R_i(x) = (2 + \frac{2}{i})d_i^2 x$ for sufficiently large $x > 0$.

**Proof.** We prove all claims by induction over $i$. The base case is $i = 2$. Observe that
\[ Q_2 = v_1^2 + v_2^2 + d_1^2 w_2^2 = 2d_1^2 w_1^2 - 2d_1 d_2 w_1 w_2 + 2d_2^2 w_2^2. \]

For fixed $w_2$, the derivative with respect to $w_1$ is
\[ \frac{\partial}{\partial w_1} Q_2 = 4d_1^2 w_1 - 2d_1 d_2 w_2, \tag{1} \]
which implies that $Q_2$ is minimized for $w_1 = \frac{d_1}{d_2} w_2$. This choice is feasible (with respect to the constraint $w_1 \geq 1$) when $w_2 \geq \frac{2d_1}{d_2}$. If $w_2 < \frac{2d_1}{d_2}$, then $\frac{\partial}{\partial w_1} Q_2$ is positive for all values of $w_1 \geq 1$, so the minimum occurs at $w_1 = 1$. It follows that
\[ Q_2(w_2) = \begin{cases} \frac{3}{4} d_1^2 w_2^2 & \text{for } w_2 \geq \frac{2d_1}{d_2}, \\
2d_2^2 w_2 - \xi_1 w_2 + 2d_1^2 & \text{otherwise}, \end{cases} \]

where $\xi_1 := 2d_1 d_2$.
and so we have
\[ R_2(w_2) = \begin{cases} 
3d^2_2 w_2 & \text{for } w_2 \geq \frac{2d_i}{d_2}, \\
4d^2_2 w_2 - \xi_1 & \text{otherwise.}
\end{cases} \] (2)

In other words, \( R_2 \) is piecewise linear and has a single breakpoint at \( \frac{2d_i}{d_2} \). The function \( R_2 \) is continuous because \( 3d^2_2 w_2 = 4d^2_2 w_2 - \xi_1 \) when \( w_2 = \frac{2d_i}{d_2} \). We have \( R_2(0) = -\xi_1 < 0 \), \( S_2(x) \geq 3d^2_2 \) for all \( x > 0 \), and \( R_2(x) = 3d^2_2 x \) for \( x \geq \frac{2d_i}{d_2} \). The fact that \( S_2(x) \geq 3d^2_2 > 0 \) makes \( R_2 \) strictly increasing.

Consider now \( i \geq 2 \), assume that \( R_i \) and \( S_i \) satisfy the induction hypothesis, and consider \( Q_{i+1} \).

By definition, we have
\[ Q_{i+1} = Q_i - \xi_i w_i w_{i+1} + 2d^2_{i+1} w_{i+1}^2. \] (3)

For a given value of \( w_{i+1} > 0 \), we need to find the value of \( w_i \) that will minimize \( Q_{i+1} \). The derivative is
\[
\frac{\partial}{\partial w_i} Q_{i+1} = R_i(w_i) - \xi_i w_{i+1}.
\]

The minimum thus occurs when \( R_i(w_i) = \xi_i w_{i+1} \).

Since \( R_i \) is a strictly increasing continuous function with \( R_i(0) < 0 \) and \( \lim_{x \to \infty} R_i(x) = \infty \), for any given \( w_{i+1} > 0 \), there exists a unique value \( w_i = R_i^{-1}(\xi_i w_{i+1}) \). However, we also need to satisfy the constraint \( w_i + w_{i+1} \geq 1 \).

We first show that \( R_{i+1} \) is continuous and piecewise linear, and that \( R_{i+1}^{-1}(0) < 0 \). We will distinguish two cases, based on the value of \( w_i := R_i^{-1}(0) \).

**Case 1:** \( w_i^0 \geq 1 \). This means that \( R_i^{-1}(\xi_i w_{i+1}) \geq 1 \) for any \( w_{i+1} \geq 0 \), and so the constraint of \( w_i + w_{i+1} \geq 1 \) is satisfied for the optimal choice of \( w_i = R_i^{-1}(\xi_i w_{i+1}) \). It follows that
\[
Q_{i+1}(w_{i+1}) = Q_i(R_i^{-1}(\xi_i w_{i+1})) - \xi_i w_{i+1} R_i^{-1}(\xi_i w_{i+1}) + 2d^2_{i+1} w_{i+1}^2.
\]
The derivative \( R_{i+1} \) is therefore
\[
R_{i+1}(w_{i+1}) = R_i(R_i^{-1}(\xi_i w_{i+1})) \frac{\xi_i}{R_i'(R_i^{-1}(\xi_i w_{i+1}))} - \xi_i R_i^{-1}(\xi_i w_{i+1}) - \xi_i w_{i+1} R_i'(R_i^{-1}(\xi_i w_{i+1})) + 4d^2_{i+1} w_{i+1} = 4d^2_{i+1} w_{i+1} - \xi_i R_i^{-1}(\xi_i w_{i+1}).
\] (4)

Since \( R_i \) is continuous and piecewise linear, so is \( R_i^{-1} \), and therefore \( R_{i+1} \) is continuous and piecewise linear. We have \( R_{i+1}(0) = -\xi_i w_i^0 < 0 \).

**Case 2:** \( w_i^0 < 1 \). Consider the function \( x \mapsto f(x) = x + R_i(x)/\xi_i \). Since \( R_i \) is continuous and strictly increasing by the inductive assumption, so is the function \( f \). Observe that \( f(w_i^0) = w_i^0 < 1 \).

As \( w_i^0 < 1 \), we have \( R_i(1) > R_i(w_i^0) = 0 \), which implies that \( f(1) > 1 \). Thus, there exists a unique value \( w_i^\infty \in (w_i^0, 1) \) such that \( f(w_i^\infty) = w_i^\infty + R_i(w_i^\infty)/\xi_i = 1 \).

For \( w_{i+1} \geq 1 - w_i^\infty = R_i(w_i^\infty)/\xi_i \), we have \( R_i^{-1}(\xi_i w_{i+1}) \geq w_i^\infty \), and so \( R_i^{-1}(\xi_i w_{i+1}) + w_{i+1} \geq 1 \). This implies that the constraint \( w_i + w_{i+1} \geq 1 \) is satisfied when \( Q_{i+1}(w_{i+1}) \) is minimized for the optimal choice of \( w_i = R_i^{-1}(\xi_i w_{i+1}) \). So \( R_{i+1} \) is as in (4) in Case 1.
When \( w_{i+1} < 1 - w_i^\infty \), the constraint \( w_i + w_{i+1} \geq 1 \) implies that \( w_i \geq 1 - w_{i+1} > w_i^\infty \). For any \( w_i > w_i^\infty \) we have \( \frac{\partial}{\partial w_i} Q_{i+1} = R_i(w_i) - \xi_i w_{i+1} > R_i(w_i^\infty) - \xi_i(1 - w_i^\infty) = 0 \). So \( Q_{i+1} \) is increasing, and the minimal value is obtained for the smallest feasible choice of \( w_i \), that is, for \( w_i = 1 - w_{i+1} \). It follows that

\[
Q_{i+1}(w_{i+1}) = Q_i(1 - w_{i+1}) - \xi_i w_{i+1}(1 - w_{i+1}) + 2d_{i+1}^2 w_{i+1}^2
\]

\[
= Q_i(1 - w_{i+1}) - \xi_i w_{i+1} + (\xi_i + 2d_{i+1}^2) w_{i+1}^2,
\]

and so the derivative \( R_{i+1} \) is

\[
R_{i+1}(w_{i+1}) = -R_i(1 - w_{i+1}) + (2\xi_i + 4d_{i+1}^2) w_{i+1} - \xi_i.
\]

Combining (4) and (5), we have

\[
R_{i+1}(w_{i+1}) = \begin{cases} 
- R_i(1 - w_{i+1}) + (2\xi_i + 4d_{i+1}^2) w_{i+1} - \xi_i & \text{for } w_{i+1} < 1 - w_i^\infty, \\
4d_{i+1}^2 w_{i+1} - \xi_i R_i^{-1}(\xi_i w_{i+1}) & \text{for } w_{i+1} \geq 1 - w_i^\infty.
\end{cases}
\]

For \( w_{i+1} = 1 - w_i^\infty \), we have \( R_i(1 - w_{i+1}) = R_i(w_i^\infty) = \xi_i(1 - w_i^\infty) \) and \( R_i^{-1}(\xi_i w_{i+1}) = R_i^{-1}(\xi_i(1 - w_i^\infty)) = w_i^\infty \), and so both expressions have the same value:

\[
-R_i(1 - w_{i+1}) + (2\xi_i + 4d_{i+1}^2) w_{i+1} - \xi_i
\]

\[
= \xi_i w_i^\infty - \xi_i + 2\xi_i - 2\xi_i w_i^\infty + 4d_{i+1}^2(1 - w_i^\infty) - \xi_i
\]

\[
= 4d_{i+1}^2(1 - w_i^\infty) - \xi_i w_i^\infty
\]

\[
= 4d_{i+1}^2(1 - w_i^\infty) - \xi_i R_i^{-1}(\xi_i w_{i+1}).
\]

Since \( R_i \) is continuous and piecewise linear, this implies that \( R_{i+1} \) is continuous and piecewise linear. We have \( R_{i+1}(0) = -R_i(1) - \xi_i \). Since \( w_i^\infty < 1 \), we have \( R_i(1) > R_i(w_i^\infty) = 0 \), and so \( R_{i+1}(0) < 0 \).

Next, we show that \( S_{i+1}(x) \geq (2 + 2/i + 1)d_{i+1}^2 \) for all \( x \geq 0 \), which implies that \( R_{i+1} \) is strictly increasing. If \( w_i^\infty < 1 \) and \( x < 1 - w_i^\infty \), then by (6),

\[
S_{i+1}(x) = S_i(1 - x) + 2\xi_i + 4d_{i+1}^2 > 4d_{i+1}^2 > (2 + 2/i + 1)d_{i+1}^2.
\]

If \( w_i^\infty \geq 1 \) or \( x > 1 - w_i^\infty \), we have by (4) and (6) that \( R_{i+1}(x) = 4d_{i+1}^2 x - \xi_i R_i^{-1}(\xi_i x) \). By the inductive assumption that \( S_i(x) \geq (2 + 2/i)d_i^2 \) for all \( x \geq 0 \), we get \( \frac{\partial}{\partial x} R_i^{-1}(x) \leq 1/(2 + 2/i)d_i^2 \). It follows that

\[
S_{i+1}(x) \geq 4d_{i+1}^2 - \frac{(2d_i d_{i+1})^2}{(2 + 2/i)d_i^2} = \left(4 - \frac{4}{2 + 2/i}\right)d_{i+1}^2 = \left(4 - \frac{2i}{i + 1}\right)d_{i+1}^2
\]

\[
= \left(2 + \frac{2i}{i + 1}\right)d_{i+1}^2.
\]

This establishes the lower bound on \( S_{i+1}(x) \).

Finally, by the inductive assumption, when \( x \) is large enough, we have \( R_i^{-1}(x) = x/(2 + 2/i)d_i^2 \), and so

\[
R_{i+1}(x) = 4d_{i+1}^2 x - \frac{(2d_i d_{i+1})^2}{(2 + 2/i)d_i^2} x = \left(2 + \frac{2i}{i + 1}\right)d_{i+1}^2 x,
\]

completing the inductive step and therefore the proof.

\[\square\]
3 The algorithm

Our algorithm progressively constructs a representation of the functions \( R_2, R_3, \ldots, R_{n-1} \). The function representation supports the following three operations:

- Op 1: given \( x \), return \( R_i(x) \);
- Op 2: given \( y \), return \( R_i^{-1}(y) \);
- Op 3: given \( \xi \), return \( x^\ast \) such that \( x^\ast + \frac{R_i(x^\ast)}{\xi} = 1 \).

The proof of Theorem 2.1 gives the relation between \( R_{i+1} \) and \( R_i \). This will allow us to construct the functions one by one—we discuss the detailed implementation in Sections 3.1 and 3.2 below.

Once all functions \( R_2, \ldots, R_{n-1} \) are constructed, the optimal weights \( w_1, w_2, \ldots, w_{n-1} \) are computed from the \( R_i \)'s as follows. Recall that \( Q = Q_{n-1} \), so \( w_{n-1} \) is the value minimizing \( Q_{n-1}(w_{n-1}) \) under the constraint \( w_{n-1} \geq 1 \). If \( R_{n-1}^{-1}(0) \geq 1 \), then \( R_{n-1}^{-1}(0) \) is the optimal value for \( w_{n-1} \); otherwise, we set \( w_{n-1} \) to 1.

To obtain \( w_{n-2} \), recall from (3) that \( Q = Q_{n-1} = Q_{n-2}(w_{n-2}) - \xi_{n-2}^2 w_{n-2} + 2d_{n-1}^2 w_{n-1}^2 \). Since we have already determined the correct value of \( w_{n-1} \), it remains to choose \( w_{n-2} \) so that \( Q_{n-1} \) is minimized. Since

\[
\frac{\partial}{\partial w_{n-2}} Q_{n-1} = R_{n-2}(w_{n-2}) - \xi_{n-2} w_{n-1},
\]

\( Q_{n-1} \) is minimized when \( R_{n-2}(w_{n-2}) = \xi_{n-2} w_{n-1} \), and so \( w_{n-2} = R_{n-2}^{-1}(\xi_{n-2} w_{n-1}) \).

In general, for \( i \in [2, n-2] \), we can obtain \( w_i \) from \( w_{i+1} \) by observing that

\[
Q_{n-1} = Q_i(w_i) - \xi_i w_i w_{i+1} + g(w_{i+1}, \ldots, w_{n-1}),
\]

where \( g \) is function that only depends on \( w_{i+1}, \ldots, w_{n-1} \). Taking the derivative again, we have

\[
\frac{\partial}{\partial w_i} Q_{n-1} = R_i(w_i) - \xi_i w_{i+1},
\]

so choosing \( w_i = R_i^{-1}(\xi_i w_{i+1}) \) minimizes \( Q_{n-1} \). To also satisfy the constraint \( w_i + w_{i+1} \geq 1 \), we need to choose \( w_i = \max\{R_i^{-1}(\xi_i w_{i+1}), 1 - w_{i+1}\} \) for \( i \in [2, n-2] \). Finally, from the discussion that immediately follows (1), we set \( w_1 = \max\{D_{2n}^{-1} w_2, 1\} \). To summarize, we have

\[
w_{n-1} = \max\{R_{n-1}^{-1}(0), 1\},
\]

\[
w_i = \max\{R_i^{-1}(\xi_i w_{i+1}), 1 - w_{i+1}\}, \quad \text{for} \ i \in [2, n-2],
\]

\[
w_1 = \max\{D_{2n}^{-1} w_2, 1\}.
\]

It follows that we can obtain the optimal weights using a single Op 2 on each \( R_i \).

3.1 Explicit representation of piecewise linear functions

Since \( R_i \) is a piecewise linear function, a natural representation is a sequence of linear functions, together with the sequence of breakpoints. Since \( R_i \) is strictly increasing, all three operations can
then be implemented to run in time $O(\log k)$ using binary search, where $k$ is the number of function pieces.

The function $R_2$ consists of exactly two pieces. We construct it directly from $d_1, d_2$, and $\xi_1$ using (2).

To construct $R_{i+1}$ from $R_i$, we first compute $w_i^0 = R_i^{-1}(0)$ using Op 2 on $R_i$. If $w_i^0 \geq 1$, then by (4) each piece of $R_i$, starting at the $x$-coordinate $w_i^0$, gives rise to a linear piece of $R_{i+1}$, so the number of pieces of $R_{i+1}$ is at most that of $R_i$.

If $w_i^0 < 1$, then we compute $w_i^\infty$ using Op 3 on $R_i$. The new function $R_{i+1}$ has a breakpoint at $1 - w_i^\infty$ by (6). Its pieces for $x \geq 1 - w_i^\infty$ are computed from the pieces of $R_i$ starting at the $x$-coordinate $w_i^\infty$. Its pieces for $0 < x < 1 - w_i^\infty$ are computed from the pieces of $R_i$ between the $x$-coordinates 1 and $w_i^\infty$. (Increasing $w_i+1$ now corresponds to a decreasing $w_i$.) This implies that every piece of $R_i$ that covers $x$-coordinates in the range $[w_i^\infty, 1]$ will give rise to two pieces of $R_{i+1}$, so the number of pieces of $R_{i+1}$ may be twice the number of pieces of $R_i$.

Therefore, although this method works, it is unclear whether the number of linear pieces of $R_i$ is bounded by a polynomial in $i$.

### 3.2 A quadratic time implementation

Since we have no polynomial bound on the number of linear pieces of the function $R_{n-1}$, we turn to an implicit representation of $R_i$.

The representation is based on the fact that there is a linear relationship between points on the graphs of the functions $R_i$ and $R_{i+1}$. Concretely, let $y_i = R_i(x_i)$, and $y_{i+1} = R_{i+1}(x_{i+1})$. Recall the following relation from (4) for the case of $w_i^0 \geq 1$:

$$R_{i+1}(w_{i+1}) = 4d_{i+1}^2w_{i+1} - \xi_iR_i^{-1}(\xi iw_{i+1}).$$

We can express this relation as a system of two equations:

$$y_{i+1} = 4d_{i+1}^2x_{i+1} - \xi_ix_i,$$
$$y_i = \xi_ix_{i+1}.$$

This can be rewritten as

$$y_{i+1} = 4d_{i+1}^2y_i/\xi_i - \xi_ix_i,$$
$$x_{i+1} = y_i/\xi_i,$$

or in matrix notation

$$\begin{pmatrix} x_{i+1} \\ y_{i+1} \\ 1 \end{pmatrix} = M_{i+1} \times \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix},$$

where

$$M_{i+1} = \begin{pmatrix} 0 & 1/\xi_i & 0 \\ -\xi_i & 4d_{i+1}^2/\xi_i & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$  \tag{7}

On the other hand, if $w_i^0 < 1$, then $R_{i+1}$ has a breakpoint at $1 - w_i^\infty$. The value $w_i^\infty$ can be obtained by applying Op 3 to $R_i$. We compute the coordinates of this breakpoint: $(1 - w_i^\infty, R_{i+1}(1 - w_i^\infty))$. Note that $R_{i+1}(1 - w_i^\infty) = 4d_{i+1}^2(1 - w_i^\infty) - \xi_iR_i^{-1}(\xi_i(1 - w_i^\infty))$ which can be computed by applying Op 2 to $R_i$. For $x_{i+1} > 1 - w_i^\infty$, the relationship between $(x_i, y_i)$ and $(x_{i+1}, y_{i+1})$ is given by (7). For $0 < x_{i+1} < 1 - w_i^\infty$, recall from (5) that

$$R_{i+1}(w_{i+1}) = -R_i(1 - w_{i+1}) + (2\xi_i + 4d_{i+1}^2)w_{i+1} - \xi_i.$$
We again rewrite this as
\[
y_{i+1} = -y_i + (2\xi_i + 4d_{i+1}^2)x_{i+1} - \xi_i,
\]
x_{i} = 1 - x_{i+1},
which gives
\[
y_{i+1} = -y_i + (2\xi_i + 4d_{i+1}^2)(1 - x_i) - \xi_i,
\]
x_{i+1} = 1 - x_i,

or in matrix notation:
\[
\begin{pmatrix}
  x_{i+1} \\
y_{i+1} \\
1
\end{pmatrix} = L_{i+1} \times \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix},
\]
where
\[
L_{i+1} = \begin{pmatrix}
  -1 & 0 & 1 \\
  -2\xi_i - 4d_{i+1}^2 & -1 & \xi_i + 4d_{i+1}^2 \\
  0 & 0 & 1
\end{pmatrix}.
\]
The function \( R_{i+1} \) is stored by storing the breakpoint \((x_{i+1}^*, y_{i+1}^*) = (1 - w_i^*, R_{i+1}(1 - w_i^*))\) as well as the two matrices \( L_{i+1} \) and \( M_{i+1} \).

Note that the first function \( R_2 \) is stored explicitly. A new function \( R_{i+1} \) can be constructed in constant time plus a constant number of queries on \( R_i \) and requires constant space only.

We now explain how the three operations Op 1, Op 2, and Op 3 are implemented on this representation of the function \( R_i \). For an operation on \( R_i \), we progressively build transformation matrices \( T_i, T_{i-1}^i, T_{i-2}^i, \ldots, T_3^i, T_2^i \) such that \((x_i, y_i, 1) = T_j^i \times (x_j, y_j, 1)\) for every \( 2 \leq j \leq i \) in a neighborhood of the query. Once we obtain \( T_j^i \), we use our explicit representation of \( R_2 \) to express \( y_i \) as a linear function of \( x_i \) in a neighborhood of the query, which then allows us to answer the query.

The first matrix \( T_i^i \) is the identity matrix. We obtain \( T_j^i \) from \( T_{j+1}^i \), for \( j \in [2, i - 1] \), as follows: If \( R_{j+1} \) has no breakpoint, then \( T_j^i = T_{j+1}^i \times M_{j+1} \). If \( R_{j+1} \) has a breakpoint \((x_{j+1}^*, y_{j+1}^*)\), then either \( T_j^i = T_{j+1}^i \times M_{j+1} \) or \( T_j^i = T_{j+1}^i \times L_{j+1} \), depending on which side of the breakpoint applies to the answer of the query. We can decide this by comparing \((x', y') = T_{j+1}^i \times (x_{j+1}^*, y_{j+1}^*+1, 1)\) with the query. More precisely, for Op 1 we compare the input \( x \) with \( x' \), for Op 2 we compare the input \( y \) with \( y' \), and for Op 3 we compute \( x' + y' / \xi \) and compare with 1.

It follows that our implicit representation of \( R_i \) supports all three operations on \( R_i \) in time \( O(i) \), and so the total time to construct \( R_{n-1} \) is \( O(n^2) \).

**Theorem 3.1.** Given \( n \) points on a line, we can compute an optimal set of weights for minimizing the quality measure \( Q \) in \( O(n^2) \) time.

## 4 Experiments

We have implemented both the explicit and implicit representations in Python. For comparison, we used the quadratic solver CVXPY\(^1\) using the modeling library PICOS\(^2\) (our code is available at [https://github.com/otfried/graph-fitting-1d](https://github.com/otfried/graph-fitting-1d)).
Table 1: Running times of the three methods (in seconds).

| $n$ | QP Explicit | Implicit |
|-----|-------------|----------|
| 100 | 0.413       | 0.00809  | 0.129    |
| 200 | 1.51        | 0.0183   | 0.353    |
| 400 | 6.38        | 0.0536   | 1.3      |
| 800 | 32.3        | 0.127    | 6.25     |
| 1600| 208         | 0.217    | 17.1     |
| 3200| 1,300       | 0.406    | 89.2     |

Table 2: Average and maximum number of pieces for three different distributions.

| $n$ | small uniform | large uniform | Gaussian |
|-----|---------------|---------------|---------|
|     | avg | max | avg | max | avg | max |
| 100 | 13.753 | 33 | 13.726 | 31 | 13.109 | 31 |
| 1000| 23.613 | 48 | 23.483 | 51 | 22.246 | 49 |
| 10000| 33.793 | 73 | 35.329 | 65 | 31.529 | 57 |
| 100000| 42.634 | 125 | 48.279 | 95 | 41.701 | 76 |

**Running times.** To compare the running time of the different methods, we first generated problem instances randomly, by setting each interpoint distance $d_i$ to an independent random value, taken uniformly from the integers $\{1, 2, \ldots, 50\}$. Table 1 shows the results.

Perhaps surprisingly, the simple method that represents each $R_i$ as a sequence of linear functions outperforms the other two methods. Apparently, at least for random interpoint distances, the number of linear pieces of these functions does not grow fast.

**Number of pieces.** To investigate this further, we have generated problem instances, with various distributions used for the random generation of interpoint distances. The results can be seen in Table 2. In the small uniform distribution, interpoint distances are taken uniformly from the set $\{1, 2, \ldots, 50\}$, for the large uniform distribution from the set $\{1, 2, \ldots, 10,000\}$. In the third column, interpoint distances are sampled from a Gaussian distribution with mean 100 and standard deviation 30. For each distribution and $n$, we compute the functions $R_2, R_3, \ldots, R_{n-1}$, and take the maximum of the number of pieces over these $n-2$ functions. We repeat each experiment 1,000 times, and show both the average and the maximum of the number of pieces found.

The table explains why the simple method performs so well in practice: as long as the number of pieces remains small, its running time is essentially linear. In fact, we are not even using binary search to implement the three operations on the piecewise linear functions.

**Precision.** The CVXOPT solver uses an iterative procedure in floating point arithmetic, and so its precision is limited. With the tolerance set to the maximum feasible value of $10^{-6}$, some weights differ from our algorithm’s solution by as much as 0.05. Our algorithm can easily be implemented using exact or high-precision arithmetic. In fact, in our implementation it suffices to provide the initial distance vector using Python Fraction objects for exact rational arithmetic, or as

1. http://cvxopt.org
2. http://picos.zib.de
high-precision floating point numbers from the `mpmath` Python library.\(^3\) Using rational arithmetic, computing the exact optimal solution for 3200 points with integer interpoint distances from the set \(\{1, 2, \ldots, 50\}\) takes between 1.4 and 4 seconds.

5 Conclusion

While in practice the explicit representation of the functions \(R_i\) works well, we do not have a polynomial time bound on the running time using this method. Future work should determine if this method can indeed be slow on some instances, or if the number of pieces can be bounded. It would also be nice to obtain an algorithm for higher dimensions that is not based on a quadratic programming solver.

In two dimensions, we have conducted some experiments that indicate that the Delaunay triangulation of the point set contains a well-fitting graph. If we choose the graph edges only from the Delaunay edges and compute the optimal edge weights, the resulting quality measure is very close to the best quality measure in the unrestricted case. It is conceivable that one can obtain a provably good approximation from the Delaunay triangulation.

References

[1] S.I. Daitch, J.A. Kelner, and D.A. Spielman. Fitting a graph to vector data. *ICML*, 2009, 201–208.

[2] S. Han, H. Huang, H. Qin, and D. Yu. Locality-preserving L1-graph and its application in clustering. *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, 2015, 813–818.

[3] T. Jebara, J. Wang, and S.-F. Chang. Graph construction and \(b\)-matching for semi-supervised learning. *ICML*, 2009, 441–448.

[4] S. Kapoor and P.M. Vaidya. Fast algorithms for convex quadratic programming and multi-commodity flows. *Proceedings of the 18th Annual ACM Symposium on Theory of Computing*, 1986, 147–159.

[5] A.Y. Ng, M.I. Jordan, and Y. Weiss. On spectral clustering: analysis and an algorithm. *NIPS*, 2001, 849–856.

[6] Y. Ye and E. Tse. An extension of Karmarkar’s projective algorithm for convex quadratic programming. *Mathematical Programming*, 44 (1989), 157–179.

[7] Y.-M. Zhang, K. Huang, and C.-L. Liu. Learning locality preserving graph from data. *IEEE Transactions on Cybernetics*, 44 (2014), 2088–2098.

[8] D. Zhou, O. Bousquet, T.N. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. *NIPS*, 2003, 321–328.

\(^3\)http://mpmath.org