Diskrete Mathematik und Optimierung

Daniel Altmann
Eugen Grycko
Winfried Hochstättler
Günther Klützke:

Monotone smoothing of noisy data

Technical Report feu-dmo034.15
Contact: daniel.math@t-online.de
eugen.grycko@fernuni-hagen.de
winfried.hochstaettler@fernuni-hagen.de
g.kluetzke@gmx.de

FernUniversität in Hagen
Fakultät für Mathematik und Informatik
Lehrgebiet für Diskrete Mathematik und Optimierung
D – 58084 Hagen

2000 Mathematics Subject Classification: 90C20,65D10,93E14
Keywords: quadratic programming, smoothing, noisy data

Monotone Smoothing of Noisy Data

Daniel Altmann, Eugen Grycko, Winfried Hochstättler
Fakultät für Mathematik und Informatik
FernUniversität in Hagen, Germany

Günter Klützke
TU Dortmund, Germany

July 19, 2014

Abstract

We consider the problem of recovering monotonicity in noisy data.

1 Introduction

Frequently one has to deal with data, where external knowledge indicates that the data should be monotone, but this property is broken due to measurement errors. We consider the problem of best approximation of given data with monotone data. Here, given data means some vector $y \in \mathbb{R}^n$, where n is large and we seek for $x \in \mathbb{R}^n$ such that $x_i \leq x_{i+1}$ for $1 \leq i \leq n-1$ and $\|x - y\|$ is minimized, where $\|\cdot\|$ is the Euclidean norm.

We present a simple algorithm with quadratic running time that solves this problem to optimality.

2 Prerequisites

Since minimizing the norm is equivalent to minimizing its square, we may as well consider

$$\min(x-y)^\top(x-y) = \min(x^\top x - 2y^\top x + y^\top y).$$

Neglecting the constant and dividing by two, in order to simplify later computations, the problem can be modelled as the following quadratic program (P)

$$\begin{array}{lll} \min & \frac{1}{2}x^\top x - y^\top x & \\ \text{subject to} & g_i(x) = x_i - x_{i+1} \quad \leq \quad 0 & \text{for all } 1 \leq i \leq n-1. \end{array} \tag{1}$$

Clearly, this program is strictly convex and hence the unique local minimum is the unique global minimum. Local minima in non-linear programming are characterized by the Karush-Kuhn-Tucker (KKT) conditions (see e.g. [1] Theorem 12.1) which we recall for later reference.

Theorem 1. *Suppose that x is a local solution of (1). Since the gradients of the constraints are linearly independent there is a Lagrange multiplier vector μ such that the following conditions are satisfied at (x, μ)*

$$\nabla f(x) + \sum_{i=1}^{n-1} \mu_i \nabla g_i(x) = 0 \tag{2}$$

$$\forall 1 \leq i \leq n-1 : g_i(x) \leq 0 \tag{3}$$

$$\forall 1 \leq i \leq n-1 : \mu_i \geq 0 \tag{4}$$

$$\forall 1 \leq i \leq n-1 : \mu_i g_i(x) = 0. \tag{5}$$

Here, the KKT-conditions are necessary and sufficient for a global minimum. In our special situation, besides feasibility the KKT conditions are equivalent to

$$x - y = \sum_{i=1}^{n-1} \mu_i (e_i - e_{i+1}) \qquad \text{(gradient equations)} \tag{6}$$

$$\mu_i (x_i - x_{i+1}) = 0 \qquad \text{for all } 1 \leq i \leq n-1 \tag{7}$$

$$\mu_i \leq 0 \qquad \text{for all } 1 \leq i \leq n-1. \tag{8}$$

The gradient equations may be interpreted as follows. We have a path with vertices e_i, the μ_i are potentials at the vertices and $x_i - y_i$ is the potential difference along edge (e_i, e_{i+1}). The simple structure of these equations allows to search for potentials which must be zero due to (7). This will split the path into two parts which can be solved independently.

This motivates the procedure Monotone_Projection in Figure 1.

(i) Initially set $x := y$.

(ii) Find an index $1 \leq k \leq n-1$ that minimizes

$$\text{diff}(k) = \frac{1}{k} \sum_{i=1}^{k} x_i - \frac{1}{n-k} \sum_{i=k+1}^{n} x_i.$$

(iii) If $\text{diff}(k) < 0$ then k is called a *splitting point*. Optimize the first k and the last $n-k$ coordinates of x recursively.

(iv) If $\text{diff}(k) \geq 0$. Set $x_j := \frac{1}{n} \sum_{i=1}^{n} x_i$ for all j and return.

Figure 1: Procedure Monotone_Projection

Clearly, the algorithm terminates after at most $n-1$ iterations. In each iteration the sums can be computed in $O(n)$. Thus, in total we have a work load of $O(n^2)$.

3 Feasibility

On termination by (iv) each y_i in a certain connected range will be replaced by the arithmetic mean of its range. By the splitting rule (iii) these ranges are disjoint and form a partition of the output vector which we denote by $\mathrm{out}(y)$

$$\mathrm{out}(y) = (\underbrace{z_1, \dots, z_1}_{\lambda_1 \text{ entries}}, \underbrace{z_2, \dots, z_2}_{\lambda_2 \text{ entries}}, \ \dots \ , \underbrace{z_h, \dots, z_h}_{\lambda_h \text{ entries}}) \tag{9}$$

where $\lambda_i \geq 1$ denotes the length of the partition and there are $h \geq 1$ different partitions. If $z_1, z_2, \dots, z_h$ is a monotone sequence, we have proven the feasibility of $\mathrm{out}(y)$. This will be our first goal.

We start with a simple observation relating the partitions alongside the splitting point to the corresponding arithmetic means. For notational convenience we use the abbreviations

$$M_k^l := \frac{1}{k}\sum_{i=1}^{k} y_i \qquad M_k^r := \frac{1}{n-k}\sum_{i=k+1}^{n} y_i,$$

for the left and right arithmetic means. We keep in mind, that the existence of a splitting point always implies $n, h \geq 2$. If k is a splitting point, we denote the left slice of y by $y[1:k]$ and the right slice by $y[k+1:n]$.

Proposition 1. *Let k be a splitting point with $k = \sum_{i=1}^{v} \lambda_i$. Let v be the largest partition index of $\mathrm{out}(y)[1:k]$ thus $v+1$ being the smallest partition index in $\mathrm{out}(y)[k+1:n]$. Recall that, since k is a splitting point we have $\mathrm{diff}(k) \leq \mathrm{diff}(\ell)$ for all $1 \leq \ell \leq n-1$. This implies that if*

i) $v \geq 2$, then

$$\frac{M_k^l - z_v}{k - \lambda_v} + \frac{M_k^r - z_v}{n - k + \lambda_v} \geq 0 \tag{10}$$

ii) $v \leq h-2$, then

$$\frac{z_{v+1} - M_k^l}{k + \lambda_{v+1}} + \frac{z_{v+1} - M_k^r}{n - k - \lambda_{v+1}} \geq 0. \tag{11}$$

Proof. i)

$$\begin{aligned} M_k^l - M_k^r &\leq M_{k-\lambda_v}^l - M_{k-\lambda_v}^r \\ &= \frac{kM_k^l - \lambda_v z_v}{k - \lambda_v} - \frac{(n-k)M_k^r + \lambda_v z_v}{n-k+\lambda_v} \\ &= M_k^l + \lambda_v \frac{M_k^l - z_v}{k - \lambda_v} - M_k^r + \lambda_v \frac{M_k^r - z_v}{n-k+\lambda_v} \\ \iff 0 &\leq \frac{M_k^l - z_v}{k-\lambda_v} + \frac{M_k^r - z_v}{n-k+\lambda_v}. \end{aligned}$$

ii)

$$\begin{aligned} M_k^l - M_k^r &\leq M_{k+\lambda_{v+1}}^l - M_{k+\lambda_{v+1}}^r \\ &= \frac{kM_k^l + \lambda_{v+1} z_{v+1}}{k + \lambda_{v+1}} - \frac{(n-k)M_k^r - \lambda_{v+1} z_{v+1}}{n-k-\lambda_{v+1}} \\ &= M_k^l + \lambda_{v+1} \frac{z_{v+1} - M_k^l}{k + \lambda_{v+1}} - M_k^r + \lambda_{v+1} \frac{z_{v+1} - M_k^r}{n-k-\lambda_{v+1}} \\ \iff 0 &\leq \frac{z_{v+1} - M_k^l}{k+\lambda_{v+1}} + \frac{z_{v+1} - M_k^r}{n-k-\lambda_{v+1}}. \end{aligned}$$

□

Lemma 1. *If* $k = \sum_{i=1}^{v} \lambda_i$ *is a splitting point then*

$$z_v < M_k^r \text{ and } z_{v+1} > M_k^l.$$

Proof. If $v = 1$ then $\text{diff}(k) < 0$ implies $M_k^l = z_1 < M_k^r$.
If $v = h-1$ then $\text{diff}(k) < 0$ implies $z_{v+1} = M_k^r > M_k^l$.

By assumption we have $M_k^l < M_k^r$, hence $z_v \geq M_k^r$ implies

$$M_k^l - z_v < M_k^r - z_v \leq 0$$

contradicting (10) for $2 \leq v \leq h-1$.

Similarly, $z_{v+1} \leq M_k^l$ implies $z_{v+1} - M_k^r < z_{v+1} - M_k^l \leq 0$ contradicting (11) for $1 \leq v \leq h-2$.

□

Corollary 1. *If* $k = \sum_{i=1}^{v} \lambda_i$ *is a splitting point then*

$$z_v < z_{v+1}.$$

Proof. Assuming $2 \leq v \leq h-2$ and multiplying (10) and (11) with their common denominators and adding the resulting equalities yields:

$$\begin{aligned} & (M_k^l - z_v)(n-k+\lambda_v) + (M_k^r - z_v)(k-\lambda_v) \\ + \quad & (z_{v+1} - M_k^l)(n-k-\lambda_{v+1}) + (z_{v+1} - M_k^r)(k+\lambda_{v+1}) \geq 0 \\ \Longleftrightarrow \quad & \underbrace{(\lambda_v + \lambda_{v+1})(M_k^l - M_k^r)}_{<0} + n(z_{v+1} - z_v) \geq 0. \end{aligned}$$

The cases $v = 1$ and $v = h-1$ remain open. We will prove only the first case, since the other one is proven in a very similar way. The first inequality holds due to the minimality of the splitting point.

$$\begin{aligned} z_v - M_k^r \leq & \frac{kz_v + \lambda_{v+1} z_{v+1}}{k + \lambda_{v+1}} - \frac{(n-k)M_k^r - \lambda_{v+1} z_{v+1}}{(n-k) - \lambda_{v+1}} \\ z_v - M_k^r \leq & z_v \frac{k}{k+\lambda_{v+1}} + z_{v+1} \left(\frac{\lambda_{v+1}}{k+\lambda_{v+1}} + \frac{\lambda_{v+1}}{(n-k)-\lambda_{v+1}} \right) \\ & - M_k^r \left(\frac{n-k}{(n-k)-\lambda_{v+1}} \right) \\ z_v \left(\frac{\lambda_{v+1}}{k+\lambda_{v+1}} \right) \leq & - M_k^r \left(\frac{\lambda_{v+1}}{(n-k)-\lambda_{v+1}} \right) \\ & + z_{v+1} \left(\frac{\lambda_{v+1} n}{(k+\lambda_{v+1})((n-k)-\lambda_{v+1})} \right) \\ z_v(n-(k+\lambda_{v+1})) - nz_{v+1} \leq & - M_k^r (k+\lambda_{v+1}) \\ nz_v - nz_{v+1} \leq & (z_v - M_k^r)(k+\lambda_{v+1}) \\ z_v - z_{v+1} \leq & \underbrace{(z_v - M_k^r)}_{<0} \frac{(k+\lambda_{v+1})}{n} < 0 \end{aligned}$$

□

Thus, $\text{out}(y)$ is feasible. Furthermore, we have proven, that according to (7) $\mu_k = 0$ must hold at every splitting point for the Lagrange multipliers of the optimum solution.

4 Optimality

Setting $\mu_0 = \mu_n = 0$ we may rewrite (6) to

$$\sum_{i=1}^{n} (\mu_i - \mu_{i-1}) e_i = x - y \tag{12}$$

and iteratively compute the values of the μ_i for $1 \leq i \leq n-1$. It is immediate that this equation is solvable if and only if

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

which is guaranteed to hold for $x = \text{out}(y)$. Hence, in order to verify the KKT-conditions for $(\text{out}(y), \mu)$ we are left to prove $\mu_i \leq 0$ for all $1 \leq i \leq n-1$.

Since the equations after a split are fully decoupled, it suffices to consider the situation that $\text{diff}(k) \geq 0$, i.e. in the terminating case (iv). Hence, we know

$$x_i = \frac{1}{n} \sum_{j=1}^{k} y_n \qquad \text{for } 1 \leq i \leq n.$$

Proposition 2. *Suppose* $\text{diff}(k) \geq 0$ *for all* $1 \leq k \leq (n-1)$*. Then setting*

$$\mu_i = \frac{i}{n} \sum_{j=1}^{n} y_j - \sum_{j=1}^{i} y_j \tag{13}$$

the vector μ *solves the gradient equation* (6)*.*

Proof. We will prove the hypothesis by induction on i.

i) $i = 1$:
The first scalar equation of (6) yields

$$\mu_1 = \frac{1}{n} \sum_{j=1}^{n} y_j - y_1.$$

ii) If (13) holds for i, then also for $i+1$ (consider equation $i+1$ of (6)):

$$\begin{aligned}
\mu_{i+1} &= \frac{1}{n} \sum_{j=1}^{n} y_j - y_{i+1} + \mu_i \\
&= \frac{1}{n} \sum_{j=1}^{n} y_i - y_{i+1} + \frac{i}{n} \sum_{j=1}^{n} y_i - \sum_{j=1}^{i} y_j \\
&= \frac{i+1}{n} \sum_{j=1}^{n} y_i - \sum_{j=1}^{i+1} y_j.
\end{aligned}$$

As mentioned above, the n-th equation is also satisfied. □

Corollary 2. *If* $\text{diff}(k) \geq 0$ *for all* $1 \leq k \leq (n-1)$*, then* (8) *and* (7) *hold for all* $1 \leq i \leq (n-1)$*.*

Proof. Condition (7) holds as $\text{out}(y)_1 = \ldots = \text{out}(y)_n = \frac{1}{n}\sum_{i=1}^{n} y_i$ and hence $g_i(x) = 0$ for $1 \leq i \leq n-1$. The following computation verifies (8):

$$\begin{aligned}
0 &\geq -\operatorname{diff}(i) \\
&= M_i^r - M_i^l \\
&= \frac{1}{n-i}\sum_{j=i+1}^{n} y_j - \frac{1}{i}\sum_{j=1}^{i} y_j \\
&= \frac{1}{i(n-i)}\left(i\sum_{j=i+1}^{n} y_j - (n-i)\sum_{j=1}^{i} y_j\right) \\
&= \frac{1}{i(n-i)}\left(i\sum_{j=1}^{n} y_j - n\sum_{j=1}^{i} y_j\right) \\
&= \frac{n}{i(n-i)}\mu_i
\end{aligned}$$

. □

Summarizing, we thus have proven.

Theorem 2. *The procedure in Figure 1 terminates with the unique* $\text{out}(y)$ *minimizing* $\|x - y\|$ *for monotone* x.

References

[1] J. Nocedal and S. J. Wright: *Numerical Optimization.* 2nd edition, Springer, New York/Berlin/Heidelberg, 2006.