

МОДИФИЦИРОВАННЫЙ MDM-АЛГОРИТМ*

А. В. Плоткин
avplotkin@gmail.com

Н. А. Соловьева
4vinyo@gmail.com

27 марта 2025 г.

Обозначим через G выпуклую оболочку некоторого конечного множества точек в n -мерном евклидовом пространстве. Рассматривается задача нахождения точки из G с наименьшей евклидовой нормой, причем начало координат (нулевой вектор) может находиться как вне, так и внутри множества G . На решение этой задачи ориентирован MDM-алгоритм [1–3]. Его сходимость в обоих случаях ($\mathbf{0} \in G$ и $\mathbf{0} \notin G$) доказана. Иногда MDM-алгоритм сходится медленно (возникает эффект «пилы»). Предпринимаются попытки уменьшить количество итераций MDM-алгоритма за счет работы с «пилами». Один из таких способов ускорения работы алгоритма описывается в [4]. В этом докладе предлагается другой способ — модифицированный вариант MDM-метода, работающий быстрее, чем принципиальный алгоритм.

1°. Пусть в пространстве \mathbb{R}^n заданы m точек,

$$H = \{a_i\}_{i=1}^m.$$

Выпуклую оболочку множества H обозначим G . Ставится экстремальная задача

$$\|v\|^2 \rightarrow \min_{v \in G}. \quad (1)$$

Очевидно, что задача (1) имеет единственное решение v^* . В случае, когда $\mathbf{0} \notin G$, v^* — ненулевой вектор. Если же $\mathbf{0} \in G$, то $v^* = \mathbf{0}$.

Обозначим через A матрицу со столбцами a_1, \dots, a_m . Тогда любой вектор v из выпуклой оболочки G множества H допускает представление $v = Ap$, где

$$p \geq \mathbf{0}, \quad \sum_{i=1}^m p[i] = 1. \quad (2)$$

*Семинар по оптимизации, машинному обучению и искусственному интеллекту «O&ML»
<http://oml.cmlaboratory.com/>

Множество векторов p , удовлетворяющих условию (2), обозначим P . Введем носитель вектора коэффициентов p — индексное множество $M_+(p)$ вида

$$M_+(p) = \{i \in 1 : m \mid p[i] > 0\}.$$

Отметим, что $M_+(p) \neq \emptyset$ для любого $p \in P$.

Величина

$$\Delta(p) = \max_{i \in M_+(p)} \langle a_i, v \rangle - \min_{i \in 1:m} \langle a_i, v \rangle,$$

где $v = Ap$, $p \in P$, называется *оценкой плана* v . Оценка плана всегда неотрицательна. Действительно,

$$\Delta(p) \geq \max_{i \in M_+(p)} \langle a_i, v \rangle - \min_{i \in M_+(p)} \langle a_i, v \rangle \geq 0.$$

В докладах [2], [3] были доказаны следующие утверждения.

ЛЕММА 1. При любом $v = Ap$, $p \in P$, справедливо неравенство

$$\|v - v^*\|^2 \leq \Delta(p).$$

ЛЕММА 2. Равенство $\Delta(p) = 0$, $p \in P$, достигается тогда и только тогда, когда вектор $v = Ap$ является решением задачи (1).

2°. Приведем описание модифицированного MDM-алгоритма решения задачи (1). Подчеркнем, что ход алгоритма будет одинаков в обоих случаях: $\mathbf{0} \notin G$ и $\mathbf{0} \in G$.

Возьмем начальное приближение $v_0 = Ap_0$, $p_0 \in P$.

Пусть уже имеется k -е приближение $v_k = Ap_k$, $p_k \in P$. Опишем k -ю итерацию алгоритма, на которой происходит построение v_{k+1} .

Обозначим

$$x_k = \max_{i \in M_+(p_k)} \langle a_i, v_k \rangle, \quad y_k = \min_{i \in 1:m} \langle a_i, v_k \rangle.$$

Оценка плана $v_k = Ap_k$ может быть записана в виде

$$\Delta_k := \Delta(p_k) = x_k - y_k.$$

Если $\Delta_k = 0$, то согласно лемме 2 вектор v_k является решением задачи (1). Процесс закончен.

Допустим, что $\Delta_k > 0$. Максимум x_k величин $\langle a_i, v_k \rangle$ по $i \in M_+(p_k)$ может достигаться на нескольких индексах одновременно. Обозначим

$$I_k^{max} = \left\{ i \in M_+(p_k) \mid \langle a_i, v_k \rangle = x_k \right\}.$$

Минимум y_k величин $\langle a_i, v_k \rangle$ по $i \in 1 : m$ может достигаться на нескольких индексах одновременно. Обозначим

$$I_k^{min} = \left\{ i \in 1 : m \mid \langle a_i, v_k \rangle = y_k \right\}.$$

Пусть

$$p'_k = \min_{i \in I_k^{max}} \{p_k[i]\}.$$

Ясно, что $p'_k > 0$. Введём вектор

$$\widehat{v}_k = v_k - p'_k \left(\sum_{i \in I_k^{max}} a_i - \frac{|I_k^{max}|}{|I_k^{min}|} \sum_{j \in I_k^{min}} a_j \right).$$

Коэффициенты разложения вектора \widehat{v}_k по столбцам матрицы A выглядят следующим образом:

$$\widehat{p}_k[i] = \begin{cases} p_k[i] & \text{при } i \notin I_k^{max}, i \notin I_k^{min}, \\ p_k[i] - p'_k & \text{при } i \in I_k^{max}, \\ p_k[i] + \frac{|I_k^{max}|}{|I_k^{min}|} p'_k & \text{при } i \in I_k^{min}. \end{cases} \quad (3)$$

Очевидно, что $\widehat{v}_k \in G$. Для простоты обозначим

$$s_k = \sum_{i \in I_k^{max}} a_i - \frac{|I_k^{max}|}{|I_k^{min}|} \sum_{j \in I_k^{min}} a_j.$$

Заметим, что

$$\frac{1}{|I_k^{max}|} s_k = \frac{1}{|I_k^{max}|} \sum_{i \in I_k^{max}} a_i - \frac{1}{|I_k^{min}|} \sum_{j \in I_k^{min}} a_j. \quad (4)$$

(В качестве направления спуска целевой функции берется вектор, соединяющий центры тяжести множеств $\{a_i\}_{i \in I_k^{max}}$ и $\{a_j\}_{j \in I_k^{min}}$.)

Покажем, что если $\Delta_k > 0$, то вектор s_k ненулевой. Предположим, напротив, что $s_k = \mathbf{0}$. Тогда из формулы (4) следует, что

$$\frac{1}{|I_k^{max}|} \sum_{i \in I_k^{max}} a_i = \frac{1}{|I_k^{min}|} \sum_{j \in I_k^{min}} a_j.$$

Умножим обе части последнего равенства на v_k , получим

$$\frac{1}{|I_k^{max}|} \sum_{i \in I_k^{max}} \langle a_i, v_k \rangle = \frac{1}{|I_k^{min}|} \sum_{j \in I_k^{min}} \langle a_j, v_k \rangle.$$

Это равносильно тому, что

$$\frac{1}{|I_k^{max}|} \sum_{i \in I_k^{max}} x_k = \frac{1}{|I_k^{min}|} \sum_{j \in I_k^{min}} y_k.$$

Последнее равенство означает, что $x_k = y_k$. Но этот факт противоречит положительности оценки $\Delta_k = x_k - y_k$.

Напомним, что

$$\widehat{v}_k = v_k - p'_k s_k.$$

Рассмотрим отрезок

$$v_k(t) = v_k + t(\widehat{v}_k - v_k) = v_k - t p'_k s_k, \quad t \in [0, 1].$$

В силу выпуклости множества G все точки $v_k(t)$ этого отрезка принадлежат G . Выберем $t_k \in [0, 1]$ из условия

$$\|v_k(t_k)\|^2 = \min_{t \in [0, 1]} \|v_k(t)\|^2.$$

Положим $v_{k+1} = v_k(t_k)$. Коэффициенты разложения вектора v_{k+1} по столбцам матрицы A выглядят так:

$$p_{k+1}[i] = \begin{cases} p_k[i] & \text{при } i \notin I_k^{max}, i \notin I_k^{min}, \\ p_k[i] - t_k p'_k & \text{при } i \in I_k^{max}, \\ p_k[i] + \frac{|I_k^{max}|}{|I_k^{min}|} t_k p'_k & \text{при } i \in I_k^{min}. \end{cases}$$

Укажем явную формулу для t_k . Запишем выражение для $\|v_k(t)\|^2$. Имеем

$$\|v_k(t)\|^2 = \|v_k - t p'_k s_k\|^2 = \|v_k\|^2 - 2t p'_k \langle v_k, s_k \rangle + t^2 (p'_k)^2 \|s_k\|^2.$$

Заметим, что

$$\begin{aligned} \langle v_k, s_k \rangle &= \left\langle v_k, \sum_{i \in I_k^{max}} a_i - \frac{|I_k^{max}|}{|I_k^{min}|} \sum_{j \in I_k^{min}} a_j \right\rangle = \\ &= \sum_{i \in I_k^{max}} \langle a_i, v_k \rangle - \frac{|I_k^{max}|}{|I_k^{min}|} \sum_{j \in I_k^{min}} \langle a_j, v_k \rangle = \\ &= |I_k^{max}| x_k - \frac{|I_k^{max}|}{|I_k^{min}|} |I_k^{min}| y_k = |I_k^{max}| \Delta_k. \end{aligned} \tag{5}$$

С учетом последней формулы получим

$$\|v_k(t)\|^2 = \|v_k\|^2 - 2t p'_k |I_k^{max}| \Delta_k + t^2 (p'_k)^2 \|s_k\|^2. \tag{6}$$

Абсолютный минимум квадратичной функции одной переменной $\|v_k(t)\|^2$ на \mathbb{R} достигается в точке

$$\widehat{t}_k = \frac{|I_k^{max}| \Delta_k}{p'_k \|s_k\|^2}. \quad (7)$$

Ясно, что $\widehat{t}_k > 0$. Значит,

$$t_k = \min \{1, \widehat{t}_k\}.$$

Если $\widehat{t}_k \leq 1$, то итерацию метода будем называть *неусеченной*, в этом случае $t_k = \widehat{t}_k$. Если $\widehat{t}_k > 1$, то итерацию метода будем называть *усеченной*, в этом случае $t_k = 1$. В силу (7) на неусеченной итерации имеем

$$v_{k+1} = v_k - \widehat{t}_k p'_k s_k = v_k - \frac{|I_k^{max}| \Delta_k}{\|s_k\|^2} s_k. \quad (8)$$

На усеченной итерации формула перехода от v_k к v_{k+1} принимает вид

$$v_{k+1} = v_k(1) = v_k - p'_k s_k, \quad (9)$$

то есть $v_{k+1} = \widehat{v}_k$.

Описание MDM-метода завершено.

Построена последовательность v_0, v_1, v_2, \dots точек из G . Если она конечна, то последний её элемент является решением задачи (1). Вообще говоря, последовательность $\{v_k\}$ бесконечна. Такая ситуация возникает, когда

$$\Delta_k > 0 \quad \text{при всех } k = 0, 1, 2, \dots \quad (10)$$

3°. Полагаем, что выполняется условие (10). Установим, что последовательность $\{\|v_k\|\}$ является строго убывающей независимо от того, принадлежит начало координат множеству G или нет.

ПРЕДЛОЖЕНИЕ 1. При всех $k = 0, 1, 2, \dots$ справедливо неравенство

$$\|v_k\|^2 - \|v_{k+1}\|^2 \geq t_k p'_k \Delta_k |I_k^{max}|. \quad (11)$$

Неравенство выполняется как равенство на неусеченных итерациях и является строгим на усеченных итерациях.

Доказательство. Пусть k -я итерация MDM-алгоритма будет неусеченной ($\widehat{t}_k \leq 1$, $t_k = \widehat{t}_k$). Тогда в силу (8) получим

$$\|v_{k+1}\|^2 = \|v_k\|^2 - 2 \frac{|I_k^{max}| \Delta_k}{\|s_k\|^2} \langle v_k, s_k \rangle + \frac{|I_k^{max}|^2 \Delta_k^2}{\|s_k\|^4} \|s_k\|^2.$$

На основании (5) запишем

$$\|v_{k+1}\|^2 = \|v_k\|^2 - 2 \frac{|I_k^{max}| \Delta_k}{\|s_k\|^2} |I_k^{max}| \Delta_k + \frac{|I_k^{max}|^2 \Delta_k^2}{\|s_k\|^2} = \|v_k\|^2 - \frac{|I_k^{max}|^2 \Delta_k^2}{\|s_k\|^2}.$$

Отсюда с учетом (7) и предположения о неусеченности итерации следует

$$\|v_k\|^2 - \|v_{k+1}\|^2 = \frac{|I_k^{max}|^2 \Delta_k^2}{\|s_k\|^2} = \widehat{t}_k p'_k \Delta_k |I_k^{max}| = t_k p'_k \Delta_k |I_k^{max}|,$$

что соответствует утверждению предложения.

В случае усеченной итерации ($\widehat{t}_k > 1$, $t_k = 1$) воспользуемся формулой (9).
Запишем

$$\|v_{k+1}\|^2 = \|v_k\|^2 - 2p'_k \langle v_k, s_k \rangle + (p'_k)^2 \|s_k\|^2.$$

Так как (см. (7))

$$p'_k \|s_k\|^2 = \frac{|I_k^{max}| \Delta_k}{\widehat{t}_k},$$

то

$$\|v_{k+1}\|^2 = \|v_k\|^2 - 2p'_k \langle v_k, s_k \rangle + p'_k \frac{|I_k^{max}| \Delta_k}{\widehat{t}_k}.$$

Отсюда с учетом (5) получим

$$\|v_{k+1}\|^2 = \|v_k\|^2 - 2p'_k |I_k^{max}| \Delta_k + p'_k \frac{|I_k^{max}| \Delta_k}{\widehat{t}_k}.$$

Имеем

$$\begin{aligned} \|v_k\|^2 - \|v_{k+1}\|^2 &= p'_k |I_k^{max}| \Delta_k + p'_k |I_k^{max}| \Delta_k - p'_k \frac{|I_k^{max}| \Delta_k}{\widehat{t}_k} = \\ &= p'_k |I_k^{max}| \Delta_k + p'_k |I_k^{max}| \Delta_k \left(1 - \frac{1}{\widehat{t}_k}\right) > p'_k |I_k^{max}| \Delta_k = t_k p'_k \Delta_k |I_k^{max}|. \end{aligned}$$

Предложение доказано. □

Из предложения 1 следует, в частности, что

$$\|v_k\| > \|v_{k+1}\| \quad \text{при всех } k = 0, 1, 2, \dots$$

Это значит, что последовательность норм $\{\|v_k\|\}$, состоящая из положительных чисел, является строго убывающей.

Приведем еще оценку уменьшения целевой функции на неусеченной k -й итерации:

$$\|v_k\|^2 - \|v_{k+1}\|^2 \geq \frac{\Delta_k^2}{d^2}, \quad (12)$$

где d — диаметр конечного множества H . Справедливость этой оценки следует из предложения 1. Действительно, с использованием (5) получим

$$\|v_k\|^2 - \|v_{k+1}\|^2 = (\hat{t}_k p'_k) \Delta_k |I_k^{max}| = \frac{|I_k^{max}|^2 \Delta_k^2}{\|s_k\|^2}. \quad (13)$$

Заметим, что

$$\frac{1}{|I_k^{max}|} s_k = \frac{1}{|I_k^{max}|} \sum_{i \in I_k^{max}} a_i - \frac{1}{|I_k^{min}|} \sum_{j \in I_k^{min}} a_j.$$

В последней формуле и уменьшаемое, и вычитаемое являются точками множества G , так как это центры тяжести двух подмножеств H : $\{a_i\}_{i \in I_k^{max}}$ и $\{a_j\}_{j \in I_k^{min}}$. Значит, норма их разности не превосходит диаметра множества H . Следовательно,

$$\frac{1}{|I_k^{max}|} \|s_k\| \leq d.$$

Очевидно, что тогда

$$\frac{|I_k^{max}|}{\|s_k\|} \geq \frac{1}{d}.$$

Приняв во внимание последнюю формулу и (13), получим требуемую оценку (12).

4°. Использование предложенной модификации позволяет бороться с зигзагообразным поведением v_k , которое нередко встречается при использовании принципиального алгоритма. Рисунки 1а–3а демонстрируют примеры таких колебаний.

На рисунках 1б–3б показаны траектории v_k при использовании модифицированного алгоритма для тех же примеров. Удастся полностью избежать зигзагообразных движений.

Однако, могут встречаться и обратные ситуации. Рисунок 4 демонстрирует пример, когда именно в модифицированном алгоритме появляются зигзагообразные колебания.

Чтобы сделать выводы об эффективности предложенной модификации, проведем численные эксперименты в пространствах большей размерности. Обозначим размерность пространства n и количество точек m , а сами точки будем выбирать случайно равномерно распределенными на кубе $[1, 10]^n$. В качестве начального приближения во всех примерах будем брать первую точку, а критерием останова будем считать выполнение неравенства $\Delta_k < 10^{-7}$.

В таблице 1 представлено среднее значение количества итераций, требуемых алгоритмам для решения 1000 сгенерированных примеров при фиксированных n и m .

Рассмотрим отдельно случай, когда начало координат принадлежит выпуклой оболочке G . На рисунках 5а и 5б представлен подобный пример. Модифицированный алгоритм сходится к началу координат за две итерации, тогда как принципиальный вариант сталкивается с зигзагообразным поведением. Чаще всего, однако, оба алгоритма испытывают сравнимые трудности, когда начало координат принадлежит выпуклой оболочке, и нельзя выбрать предпочтительный вариант. Для демонстрации данного факта, проведем численные эксперименты в пространствах большей размерности. Точки будем выбирать случайно равномерно распределенными на кубе $[-10, 10]^n$. При достаточно большом количестве точек, выбранных указанным способом, нулевой вектор будет находиться внутри выпуклой оболочки. Как и раньше, критерием останковки будем считать выполнение неравенства $\Delta_k < 10^{-7}$.

В таблице 2 представлено среднее значение количества итераций, требуемых алгоритмам для решения 1000 сгенерированных примеров при фиксированных n и m . Можно увидеть, что полученные значения для вариантов алгоритмов практически не отличаются. При этом оба алгоритма делают значительно больше итераций, чем в случае, когда начало координат находится вне выпуклой оболочки G .

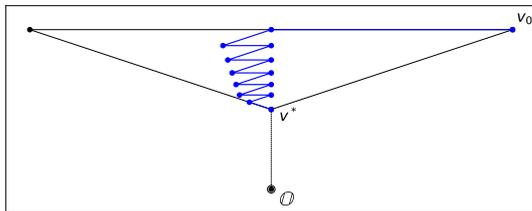


Рис. 1а. Принципиальный алгоритм
(13 шагов)

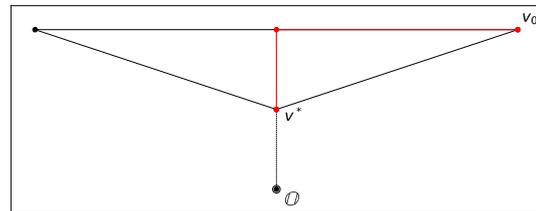


Рис. 1б. Модифицированный
алгоритм (2 шага)

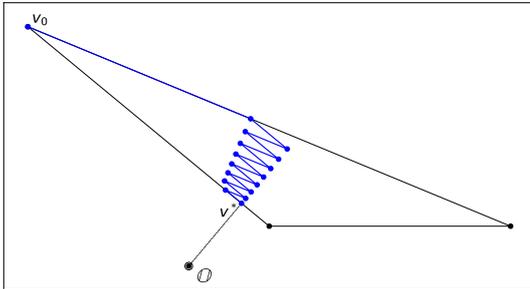


Рис. 2а. Принципиальный алгоритм
(16 шагов)

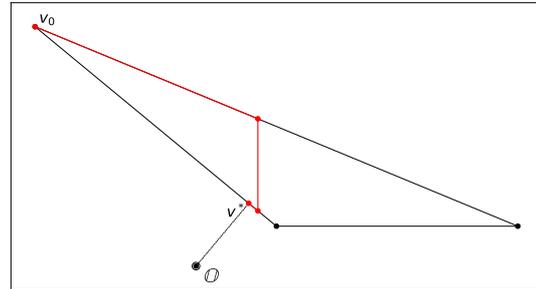


Рис. 2б. Модифицированный алгоритм (3 шага)

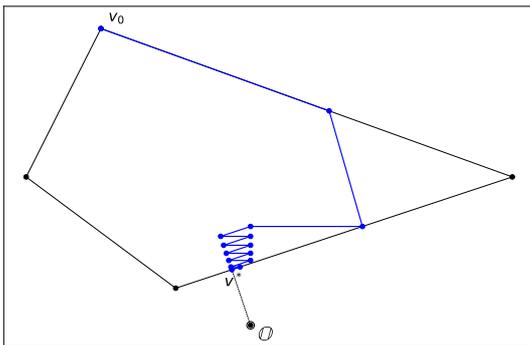


Рис. 3а. Принципиальный алгоритм
(14 шагов)

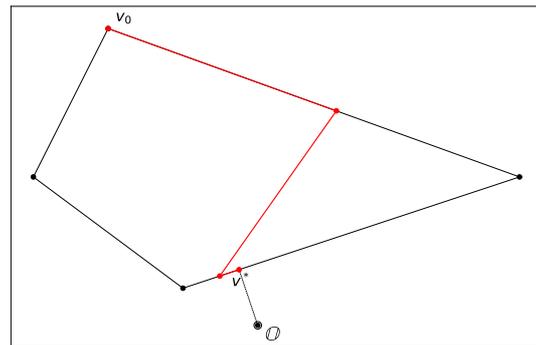


Рис. 3б. Модифицированный алгоритм (3 шага)

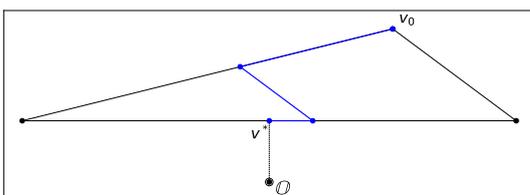


Рис. 4а. Принципиальный алгоритм
(3 шага)

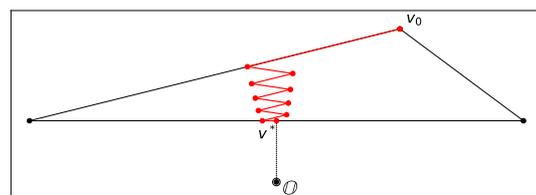


Рис. 4б. Модифицированный алгоритм (10 шагов)

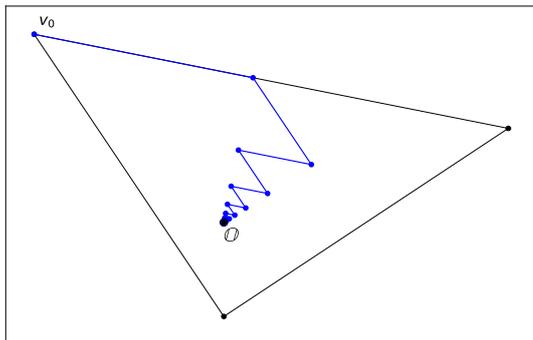


Рис. 5а. Принципиальный алгоритм
(53 шага)

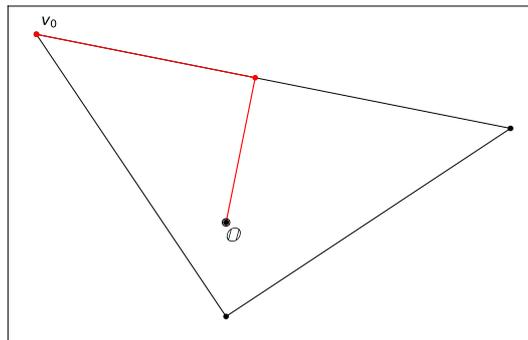


Рис. 5б. Модифицированный алгоритм
(2 шага)

Таблица 1. Среднее количество итераций для различных n и m
в случае, когда $\mathbb{O} \notin G$

n	m	Принципиальный алгоритм	Модифицированный алгоритм
5	100	24.4	16.9
10	100	42.2	31.8
20	100	45.5	37.5
30	300	55.8	50.5
50	1000	73.3	68.2

Таблица 2. Среднее количество итераций для различных n и m
в случае, когда $\mathbb{O} \in G$

n	m	Принципиальный алгоритм	Модифицированный алгоритм
5	100	45.2	45.1
10	100	111.6	111.7
20	100	394.2	392.2
30	300	327.4	330.1
50	1000	362.3	359.6

ЛИТЕРАТУРА

1. Митчелл Б. Ф., Демьянов В. Ф., Малоземов В. Н. *Нахождение ближайшей к началу координат точки многогранника* // Вестник ЛГУ. 1971. № 19. С. 38–45.
2. Малоземов В. Н. *МДМ-методу — 50 лет* // Семинар «ОМЛ». Литература. Публикации участников семинара.
(<http://oml.cmlaboratory.com/literatura.shtml>)
3. Малоземов В. Н., Плоткин А. В., Соловьева Н. А. *Универсальность МДМ-алгоритма* // Семинар «ОМЛ». Избранные доклады. 27 февраля 2025 г.
(<http://oml.cmlaboratory.com/rep25.shtml#0225>)
4. Barbero A., Lopez J., Dorronsoro J. R. *An accelerated MDM algorithm for SVM training* / European Symposium on Artificial Neural Networks — Advances in Computational Intelligence and Learning. Bruges (Belgium). 23–25 April 2008. Proceedings, pp. 421–426.