DISTANCES

Dinh Huu Nguyen, 04/13/2020

Abstract: an exposition on distances.

Contents

1. Distances

Consider dataset X of n samples $x_1, \dots, x_n$ and m features $X_1, \dots, X_m$

X	X_1	$\dots$	X_m
x_1	x_{11}	$\dots$	x_{1m}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
x_n	x_{n1}	$\dots$	x_{nm}

We view $x_1, \dots, x_n$ as vectors in $\mathbb{R}^m$ and $X_1, \dots, X_m$ as vectors in $\mathbb{R}^n$. We also write X as a matrix

$$X_{n\times m} = \begin{pmatrix} x_{11} & \dots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nm} \end{pmatrix}$$

1.1. **Euclidean distance.** The Euclidean norm and Euclidean distance on $\mathbb{R}^m$ are independent of X.

Definition 1.1. We define the Euclidean norm for any vector $u \in \mathbb{R}^m$ as

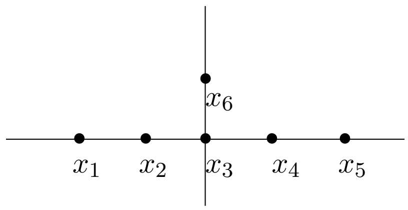
$$||u||_E = \sqrt{u^t u}$$

Definition 1.2. We define the Euclidean distance between any two vectors $u, u' \in \mathbb{R}^m$ as

$$d_E(u, u') = ||u - u'||_E$$

Example 1.3. Consider the following dataset

X	X_1	X_2
x_1	-2	0
x_2	-1	0
x_3	0	0
x_4	1	0
x_5	2	0
x_6	0	1



Then

$$\begin{aligned}
\|(1,0)\|_E^2 &= \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
&= 1 \\
\|(0,1)\|_E^2 &= \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
&= 1 \\
d_E^2(x_4, x_3) &= \|x_4 - x_3\|_E^2 \\
&= \|(1,0)\|_E^2 \\
&= 1 \\
d_E^2(x_6, x_3) &= \|x_6 - x_3\|_E^2 \\
&= \|(0,1)\|_E^2 \\
&= 1
\end{aligned}$$

1.2. **Mahalanobis distance.** The Mahalanobis norm and Mahalanobis distance on $\mathbb{R}^m$ are dependent of X. To define them X must have invertible covariance matrix

$$\begin{aligned}
M^{-1} &= X^t X \\
&= \begin{pmatrix} cov(X_1, X_1) & \dots & cov(X_1, X_m) \\ \vdots & \ddots & \vdots \\ cov(X_m, X_1) & \dots & cov(X_m, X_m) \end{pmatrix}
\end{aligned}$$

Being a Gram matrix means M^{-1} is positive semidefinite. And being invertible means it is positive definite.

Definition 1.4. Suppose X has invertible covariance matrix M^{-1}. We define the Mahalanobis norm for any vector $u \in \mathbb{R}^m$ as

$$\|u\|_M = \sqrt{u^t M u}$$

Definition 1.5. Suppose X has invertible covariance matrix M^{-1}. We define the Mahalanobis distance between any two vectors $u, u' \in \mathbb{R}^m$ as

$$d_M(u, u') = \|u - u'\|_M$$

Example 1.6. Consider the same dataset X in example 1.3. It has invertible covariance matrix

$$M^{-1} = \begin{pmatrix} \frac{5}{3} & 0 \\ 0 & \frac{5}{36} \end{pmatrix}$$

with inverse

$$M = \begin{pmatrix} \frac{3}{5} & 0 \\ 0 & \frac{36}{5} \end{pmatrix}$$

Then

$$\begin{aligned}
\|(1,0)\|_M^2 &= \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{3}{5} & 0 \\ 0 & \frac{36}{5} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
&= \frac{3}{5} \\
\|(0,1)\|_M^2 &= \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{3}{5} & 0 \\ 0 & \frac{36}{5} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
&= \frac{36}{5} \\
d_M^2(x_4, x_3) &= \|x_4 - x_3\|_M^2 \\
&= \|(1,0)\|_M^2 \\
&= \frac{3}{5} \\
d_M^2(x_6, x_3) &= \|x_6 - x_3\|_M^2 \\
&= \|(0,1)\|_M^2 \\
&= \frac{36}{5}
\end{aligned}$$

Along $(1,0)$ Mahalanobis norm/distance squared equals Euclidean norm/distance squared scaled by $\frac{3}{5}$.
Along $(0,1)$ Mahalanobis norm/distance squared equals Euclidean norm/distance squared scaled by $\frac{36}{5}$.

More generally, if $X_1, \ldots, X_m$ are uncorrelated or independent then the covariance matrix is

$$M^{-1} = \begin{pmatrix} cov(X_1, X_1) & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & cov(X_m, X_m) \end{pmatrix}$$

with inverse

$$M = \begin{pmatrix} \frac{1}{cov(X_1, X_1)} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \frac{1}{cov(X_m, X_m)} \end{pmatrix}$$

and along principal component v_i Mahalanobis norm/distance squared equals Euclidean norm/distance squared scaled by the reciprocal of corresponding eigenvalue $cov(X_i, X_i)$. And if $X_1, \ldots, X_m$

have been normalized then the covariance matrix is

$$M^{-1} = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix}$$

with inverse

$$M = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix}$$

and along principal component v_i Mahalanobis norm/distance squared equals Euclidean norm/distance squared.

Most generally, if $X_1, \dots, X_m$ are correlated then the covariance matrix M^{-1} is full of nonzero entries but the same scaling phenomenon still holds.

Proposition 1.7. *Consider dataset X of n samples $x_1, \dots, x_n$ and m features $X_1, \dots, X_m$. Suppose X has invertible covariance matrix M^{-1} with eigenvalue decomposition*

$$\begin{aligned} M^{-1} &= V \Lambda V^t \\ &= \begin{pmatrix} v_1 & \dots & v_m \end{pmatrix} \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_m \end{pmatrix} \begin{pmatrix} v_1^t \\ \vdots \\ v_m^t \end{pmatrix} \end{aligned}$$

Let $u = \alpha_1 v_1 + \dots + \alpha_m v_m$ and $u' = \beta v_1 + \dots + \beta v_m$ be two vectors in $\mathbb{R}^m$. Then

$$u^t M u' = \frac{1}{\lambda_1} \alpha_1 \beta_1 + \dots + \frac{1}{\lambda_m} \alpha_m \beta_m$$

Proof. This follows straight from definition

$$\begin{aligned} u^t M u' &= (\alpha_1 v_1 + \dots + \alpha_m v_m)^t \begin{pmatrix} v_1 & \dots & v_m \end{pmatrix} \begin{pmatrix} \frac{1}{\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\lambda_m} \end{pmatrix} \begin{pmatrix} v_1^t \\ \vdots \\ v_m^t \end{pmatrix} (\beta_1 v_1 + \dots + \beta_m v_m) \\ &= (\alpha_1 v_1^t + \dots + \alpha_m v_m^t) \begin{pmatrix} v_1 & \dots & v_m \end{pmatrix} \begin{pmatrix} \frac{1}{\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\lambda_m} \end{pmatrix} \begin{pmatrix} v_1^t \\ \vdots \\ v_m^t \end{pmatrix} (\beta_1 v_1 + \dots + \beta_m v_m) \\ &= \begin{pmatrix} \alpha_1 & \dots & \alpha_m \end{pmatrix} \begin{pmatrix} \frac{1}{\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\lambda_m} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} \\ &= \frac{1}{\lambda_1} \alpha_1 \beta_1 + \dots + \frac{1}{\lambda_m} \alpha_m \beta_m \end{aligned}$$

□

Corollary 1.8. *Consider the same dataset X in proposition 1.7. Let $u = \alpha_1 v_1 + \dots + \alpha_m v_m$ be any vector in $\mathbb{R}^m$. Then*

$$||u||_M^2 = \frac{1}{\lambda_1} \alpha_1^2 + \dots + \frac{1}{\lambda_m} \alpha_m^2$$

Proof. This follows straight from proposition 1.7. □

Corollary 1.9. *Consider the same dataset X in proposition 1.7. Let u be any vector in $\mathbb{R}^m$.*

a. If u lies along principal component v_i then

$$\|u\|_M^2 = \frac{1}{\lambda_i}\|u\|_E^2$$

b. If $u = u_1 + \cdots + u_m$ where u_i lies along principal component v_i then

$$\|u\|_M^2 = \frac{1}{\lambda_1}\|u_1\|_E^2 + \cdots + \frac{1}{\lambda_m}\|u_m\|_E^2$$

Proof. If $u = \alpha_i v_i$ then

$$\begin{aligned}\|u\|_M^2 &= \frac{1}{\lambda_i}\alpha_i^2 && \text{(corollary 1.8)}\\ &= \frac{1}{\lambda_i}\|u\|_E^2\end{aligned}$$

If $u = u_1 + \cdots + u_m$ where u_i lies along principal component v_i then

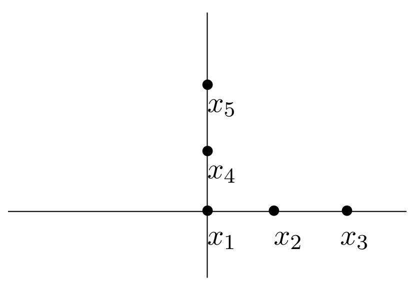
$$\begin{aligned}\|u\|_M^2 &= \|u_1 + \cdots + u_m\|_M^2\\ &= \|u_1\|_M^2 + \cdots + \|u_m\|_M^2 && \text{(corollary 3.9)}\\ &= \frac{1}{\lambda_1}\|u_1\|_E^2 + \cdots + \frac{1}{\lambda_m}\|u_m\|_E^2 && \text{(part a)}\end{aligned}$$

□

Without squaring, along principal component v_i Mahalanobis norm/distance equals Euclidean norm/distance scaled by the reciprocal of the corresponding singular value σ_i.

Example 1.10. Consider the following dataset

X	X_1	X_2
x_1	0	0
x_2	1	0
x_3	2	0
x_4	0	1
x_5	0	2



It has invertible covariance matrix

$$M^{-1} = \begin{pmatrix} \frac{16}{25} & -\frac{9}{25} \\ -\frac{9}{25} & \frac{16}{25} \end{pmatrix}$$

with inverse

$$M = \begin{pmatrix} \frac{16}{7} & \frac{9}{7} \\ \frac{9}{7} & \frac{16}{7} \end{pmatrix}$$

and eigenvalue decomposition

$$M^{-1} = \begin{pmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{7}{25} \end{pmatrix} \begin{pmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$$

and principal components

$$v_1 = \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right), v_2 = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$$

and corresponding eigenvalues

$$\lambda_1 = 1, \lambda_2 = \frac{7}{25}$$

Then $(1,0)$ does not lie along any principal component. We can use the definition of Mahalanobis norm

$$\begin{aligned} \|(1,0)\|_M^2 &= \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{16}{7} & \frac{9}{7} \\ \frac{9}{7} & \frac{16}{7} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \frac{16}{7} \end{aligned}$$

Or we can write $(1,0) = -\frac{\sqrt{2}}{2}v_1 + \frac{\sqrt{2}}{2}v_2$. Then

$$\begin{aligned} \|(1,0)\|_M^2 &= \left\|-\frac{\sqrt{2}}{2}v_1 + \frac{\sqrt{2}}{2}v_2\right\|_M^2 \\ &= \left\|-\frac{\sqrt{2}}{2}v_1\right\|_M^2 + \left\|\frac{\sqrt{2}}{2}v_2\right\|_M^2 && \text{(corollary 3.9)} \\ &= \frac{1}{2}\|v_1\|_M^2 + \frac{1}{2}\|v_2\|_M^2 \\ &= \frac{1}{2} \cdot \frac{1}{\lambda_1}\|v_1\|_E^2 + \frac{1}{2} \cdot \frac{1}{\lambda_2}\|v_2\|_E^2 && \text{(corollary 1.9 a)} \\ &= \frac{1}{2} \cdot \frac{1}{1} \cdot 1 + \frac{1}{2} \cdot \frac{25}{7} \cdot 1 \\ &= \frac{16}{7} \end{aligned}$$

2. Distances after linear dimension reduction

In linear dimension reduction, we use a matrix

$$U_{m\times l} = \begin{pmatrix} u_{11} & \dots & u_{1l} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{ml} \end{pmatrix}$$

to reduce dimension of the feature space from m to l

$$\mathbb{R}^m \xrightarrow{U^t} \mathbb{R}^l$$

$$x_j \mapsto U^t x_j = x'_j$$

Doing this for all x_j can be written as $U^t_{l\times m} X^t_{m\times n} = X'^t_{l\times n}$. Now the transformed dataset is

X'	X'_1	$\dots$	X'_l
x'_1	x'_{11}	$\dots$	x'_{1l}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
x'_n	x'_{n1}	$\dots$	x'_{nl}

2.1. **Euclidean distance.** Let us look at Euclidean distance before and Euclidean distance after reduction.

Before reduction

$$d_E^2(x_j, x_{j'}) = (x_j - x_{j'})^t (x_j - x_{j'})$$

After reduction

$$\begin{aligned} d_E^2(U^t x_j, U^t x_{j'}) &= (U^t x_j - U^t x_{j'})^t (U^t x_j - U^t x_{j'}) \\ &= (U^t (x_j - x_{j'}))^t (U^t (x_j - x_{j'})) \\ &= (x_j - x_{j'})^t U U^t (x_j - x_{j'}) \end{aligned}$$

The difference is

$$\begin{aligned} D_E^2(x_j, x_{j'}) &= d_E^2(x_j, x_{j'}) - d_E^2(U^t x_j, U^t x_{j'}) \\ &= (x_j - x_{j'})^t (I - UU^t)(x_j - x_{j'}) \end{aligned}$$

People have investigated $\mu(D_E^2), var(D_E^2)$ in terms of properties of U, whose entries are random variables U_{ih} for $1 \le i \le m, 1 \le h \le l$.

Example 2.1. Choosing $x_j - x_{j'} = (0, \dots, 1_i, \dots, 0)$ tells us that

$$d_E^2(x_j, x_{j'}) = 1$$

while

$$\begin{aligned} d_E^2(U^t x_j, U^t x_{j'}) &= \begin{pmatrix} 0 & \dots & 1_i & \dots & 0 \end{pmatrix} UU^t \begin{pmatrix} 0 \\ \vdots \\ 1_i \\ \vdots \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} u_{i1} & \dots & u_{il} \end{pmatrix} \begin{pmatrix} u_{i1} \\ \vdots \\ u_{il} \end{pmatrix} \\ &= \|u_{i*}\|_E^2 \end{aligned}$$

Preservation of distance by U means at least its rows u_{i*} must be normal $1 = \|u_{i*}\|_E^2$.

2.2. **Mahalanobis distance.** Let us look at Mahalanobis distance before and Mahalanobis distance after reduction. Suppose X has invertible covariance matrix

$$M^{-1} = X^t X$$

and X' has invertible covariance matrix

$$\begin{aligned} M'^{-1} &= X'^t X' \\ &= (XU)^t XU \\ &= U^t X^t X U \\ &= U^t M^{-1} U \end{aligned}$$

Before reduction

$$d^2_M(x_j, x_{j'}) = (x_j - x_{j'})^t M (x_j - x_{j'})$$

After reduction

$$\begin{aligned} d^2_{M'}(U^t x_j, U^t x_{j'}) &= (U^t x_j - U^t x_{j'})^t M' (U^t x_j - U^t x_{j'}) \\ &= (U^t (x_j - x_{j'}))^t M' (U^t (x_j - x_{j'})) \\ &= (x_j - x_{j'})^t U M' U^t (x_j - x_{j'}) \end{aligned}$$

The difference is

$$\begin{aligned} D^2_M(x_j, x_{j'}) &= d^2_M(x_j, x_{j'}) - d^2_{M'}(U^t x_j, U^t x_{j'}) \\ &= (x_j - x_{j'})^t (M - U M' U^t)(x_j - x_{j'}) \end{aligned}$$

We may investigate $\mu(D^2_M), var(D^2_M)$ in terms of properties of U and properties of X, based on works for the Euclidean case and results in subsection 1.2.

Example 2.2. If U has a right inverse U^{-1} such that $UU^{-1} = I$ then

$$\begin{aligned} d^2_{M'}(U^t x_j, U^t x_{j'}) &= (x_j - x_{j'})^t U M' U^t (x_j - x_{j'}) \\ &= (x_j - x_{j'})^t U U^{-1} M (U^t)^{-1} U^t (x_j - x_{j'}) \\ &= (x_j - x_{j'})^t M (x_j - x_{j'}) \end{aligned}$$

The difference is

$$\begin{aligned} D^2_M(x_j, x_{j'}) &= d^2_M(x_j, x_{j'}) - d^2_{M'}(U^t x_j, U^t x_{j'}) \\ &= 0 \end{aligned}$$

and Mahalanobis distance before and Mahalanobis distance after are the same.

However, to have right inverse U must be square or wide, or $m \le l$, which is never the case in dimension reduction.

3. Mahalanobis Inner Product

We can define the Mahalanobis inner product that induces the Mahalanobis norm and the Mahalanobis distance in section 1 plus Mahalanobis orthogonality, Mahalanobis correlation coefficient, Mahalanobis angle.

Consider a $m \times m$ positive definite matrix M^{-1} with eigenvalue decomposition

$$\begin{aligned} M^{-1} &= V \Lambda V^t \\ &= \begin{pmatrix} v_1 & \dots & v_m \end{pmatrix} \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_m \end{pmatrix} \begin{pmatrix} v_1^t \\ \vdots \\ v_m^t \end{pmatrix} \end{aligned}$$

with principal components $v_1, \dots, v_m$ and corresponding positive eigenvalues $\lambda_1, \dots, \lambda_m$.

Definition 3.1. We define the Mahalanobis inner product on $\mathbb{R}^m$ as

$$\langle u, u' \rangle_M = u^t M u'$$

If one does not want to work with coordinates and matrices, one can use the principal components as basis vectors.

Definition 3.2. We define the Mahalanobis inner product on $\mathbb{R}^m$ by defining it for basis vectors

$$\langle v_i, v_j \rangle_M = \begin{cases} \frac{1}{\lambda_i} \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

and extending that linearly

$$\begin{aligned} \langle \alpha_1 v_1 + \cdots + \alpha_m v_m, \beta_1 v_1 + \cdots + \beta_m v_m \rangle_M &= \sum_{i,j} \langle \alpha_i v_i, \beta_j v_j \rangle \\ &= \sum_i \alpha_i \beta_i \langle v_i, v_i \rangle \\ &= \sum_i \alpha_i \beta_i \frac{1}{\lambda_i} \end{aligned}$$

These two definitions agree by proposition 1.7. One can verify that this Mahalanobis inner product satisfies the three axioms symmetry, linearity and positive definiteness of inner product. It induces the Mahalanobis norm in definition 1.4, the Mahalanobis distance in definition 1.5 and more.

Definition 3.3. We say two vectors $u, u' \in \mathbb{R}^m$ is Mahalanobis orthogonal, denoted $u \perp_M u'$ if $\langle u, u' \rangle_M = 0$.

Definition 3.4. We define the Mahalanobis correlation coefficient between any two nonzero vectors $u, u' \in \mathbb{R}^m$ as

$$\rho_M(u, u') = \frac{\langle u, u' \rangle_M}{||u||_M ||u'||_M}$$

Definition 3.5. We define the Mahalanobis angle between any two nonzero vectors $u, u' \in \mathbb{R}^m$ as

$$\angle_M(u, u') = \arccos(\rho_M(u, u'))$$

Example 3.6. By definition 3.2 $\langle v_i, v_j \rangle_M = 0$. So any two principal components v_i, v_j are Mahalanobis orthogonal, their Mahalanobis correlation coefficient is 0 and their Mahalanobis angle is 90°.

Example 3.7. While orthogonal principal components are also Mahalanobis orthogonal, the same may not hold for other vectors $u, u' \in \mathbb{R}^m$. Write $u = \alpha_1 v_1 + \cdots + \alpha_m v_m$ and $u' = \beta_1 v_1 + \cdots + \beta_m v_m$. Then

$$\begin{aligned} \langle u, u' \rangle &= \langle \alpha_1 v_1 + \cdots + \alpha_m v_m, \beta_1 v_1 + \cdots + \beta_m v_m \rangle \\ &= \sum_i \alpha_i \beta_i \end{aligned}$$

while

$$\begin{aligned} \langle u, u' \rangle_M &= \langle \alpha_1 v_1 + \cdots + \alpha_m v_m, \beta_1 v_1 + \cdots + \beta_m v_m \rangle_M \\ &= \sum_i \alpha_i \beta_i \frac{1}{\lambda_i} \end{aligned}$$

Hence $\langle u, u' \rangle$ may be 0 while $\langle u, u' \rangle_M$ may not be zero or vice versa.

Some results for Euclidean notions also hold for their Mahalanobis analogues.

Theorem 3.8. *(Pythagorean theorem) If two vectors $u, u' \in \mathbb{R}^m$ are Mahalanobis orthogonal then $||u + u'||_M^2 = ||u||_M^2 + ||u'||_M^2$.*

Proof. This follows from the properties of inner product

$$\begin{aligned}
||u + u'||_M^2 &= \langle u + u', u + u' \rangle_M \\
&= \langle u, u \rangle_M + \langle u, u' \rangle_M + \langle u', u \rangle_M + \langle u', u' \rangle_M \\
&= ||u||_M^2 + 0 + 0 + ||u'||_M^2 \\
&= ||u||_M^2 + ||u'||_M^2
\end{aligned}$$

□

Corollary 3.9. *If two vectors $u, u' \in \mathbb{R}^m$ lie along principal components v, v' then $||u + u'||_M^2 = ||u||_M^2 + ||u'||_M^2$.*

Proof. This follows from the fact $v \perp_M v'$ in example 3.6 and theorem 3.8. □