Ch. 12
Linear Bayesian Estimators

Introduction

In chapter 11 we saw:

the MMSE estimator takes a simple form when $\mathbf{x}$ and $\boldsymbol{\theta}$ are jointly Gaussian – it is linear and used only the 1st and 2nd order moments (means and covariances).

Without the Gaussian assumption, the General MMSE estimator requires integrations to implement – undesirable!

So what to do if we can't "assume Gaussian" but want MMSE?

Keep the MMSE criteria

But…restrict the form of the estimator to be *LINEAR*

$\Rightarrow$ "LMMSE Estimator"

Something similar to BLUE!

LMMSE Estimator = "Wiener Filter"

Bayesian Approaches

MMSE
"Squared" Cost Function
(Nonlinear Estimate)

Estimate: $\hat{\boldsymbol{\theta}} = E\left\{\boldsymbol{\theta} \middle| \mathbf{x}\right\}$

Err. Cov.: $\mathbf{M}_{\hat{\boldsymbol{\theta}}} = E_{\mathbf{x}}\left\{\mathbf{C}_{\boldsymbol{\theta}|\mathbf{x}}\right\}$

Jointly Gaussian **x** and **θ**
(Yields Linear Estimate)

Estimate: $\hat{\boldsymbol{\theta}} = E\{\boldsymbol{\theta}\} + \mathbf{C}_{\boldsymbol{\theta}\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}\left(\mathbf{x} - E\{\mathbf{x}\}\right)$

Err. Cov.: $\mathbf{M}_{\hat{\boldsymbol{\theta}}} = \mathbf{C}_{\boldsymbol{\theta\theta}} - \mathbf{C}_{\boldsymbol{\theta}\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{C}_{\mathbf{x}\boldsymbol{\theta}}$

Bayesian Linear Model
(Yields Linear Estimate)

Estimate: $\hat{\boldsymbol{\theta}} = E\{\boldsymbol{\theta}\} + \mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T + \mathbf{C}_w\right)^{-1}\left(\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{\theta}}\right)$

Err. Cov.: $\mathbf{M}_{\hat{\boldsymbol{\theta}}} = \mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T + \mathbf{C}_w\right)^{-1}\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}$

LMMSE
Force Linear Estimate
Known: E{**θ**},E{**x**}, C

Estimate: $\hat{\boldsymbol{\theta}} = E\{\boldsymbol{\theta}\} + \mathbf{C}_{\boldsymbol{\theta}\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}\left(\mathbf{x} - E\{\mathbf{x}\}\right)$

Err. Cov.: $\mathbf{M}_{\hat{\boldsymbol{\theta}}} = \mathbf{C}_{\boldsymbol{\theta\theta}} - \mathbf{C}_{\boldsymbol{\theta}\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{C}_{\mathbf{x}\boldsymbol{\theta}}$

Same!

MAP
"Hit-or-Miss"
Cost Function

Other Cost Functions

12.3 Linear MMSE Estimator Solution

Scalar Parameter Case:

Estimate: θ, a random variable realization

Given: data vector $\mathbf{x} = [x[0]\ x[1]\ ...x[N\text{-}1]\ ]^T$

Assume:

- Joint PDF $p(\mathbf{x}, \theta)$ is unknown
- But…its 1st two moments *are* known
- There is some statistical dependence between **x** and θ
  - E.g., Could estimate θ = salary using **x** = 10 past years' taxes owed
  - E.g., Can't estimate θ = salary using **x** = 10 past years' number of Christmas cards sent

Goal: Make the best possible estimate while using an affine form for the estimator

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n] + a_N$$

Handles Non-Zero Mean Case

Choose $\{a_n\}$ to minimize $Bmse(\hat{\theta}) = E_{\mathbf{x}\theta}\{(\theta - \hat{\theta})^2\}$

Derivation of Optimal LMMSE Coefficients

Using the desired affine form of the estimator, the Bmse is

$$Bmse(\hat{\theta}) = E\left\{\left[\theta - \sum_{n=0}^{N-1} a_n x[n] + a_N\right]^2\right\}$$

Step #1: Focus on a_N
$$\frac{\partial Bmse(\hat{\theta})}{\partial a_N} = 0$$

Passing $\partial/\partial a_N$ through $E\{\}$ gives
$$-2E\{\theta - \sum_{n=0}^{N-1} a_n x[n] + a_N\} = 0$$

$$\Rightarrow \quad \boxed{a_N = E\{\theta\} - \sum_{n=0}^{N-1} a_n E\{x[n]\}}$$

Note: $a_N = 0$ if $E\{\theta\} = E\{x[n]\} = 0$

Step #2: Plug-In Step #1 Result for a_N

$$Bmse(\hat{\theta}) = E\left\{\left[\sum_{n=0}^{N-1} a_n (x[n] - E\{x[n]\}) \; - \; (\theta - E\{\theta\})\right]^2\right\}$$

$$= E\left\{\left[\underbrace{\mathbf{a}^T (\mathbf{x} - E\{\mathbf{x}\})}_{scalar} - \underbrace{(\theta - E\{\theta\})}_{scalar}\right]^2\right\}$$

where $\mathbf{a} = [a_0 \; a_1 \; . \; . \; . \; a_{N\text{-}1}]^{\mathrm{T}}$

Only up to N-1

Note: $\mathbf{a}^{\mathrm{T}} (\mathbf{x} - E\{\mathbf{x}\}) = (\mathbf{x} - E\{\mathbf{x}\})^T \mathbf{a}$ since it is scalar

Thus, expanding out $[\mathbf{a}^{\mathrm{T}}(\mathbf{x} - E\{\mathbf{x}\}) - (\boldsymbol{\theta} - E\{\boldsymbol{\theta}\})]^2$ gives

$$Bmse(\hat{\theta}) = E\left\{\mathbf{a}^T(\mathbf{x} - E\{\mathbf{x}\})(\mathbf{x} - E\{\mathbf{x}\})^T\mathbf{a}\right\} + Etc.$$

$$= \mathbf{a}^T E\left\{(\mathbf{x} - E\{\mathbf{x}\})(\mathbf{x} - E\{\mathbf{x}\})^T\right\}\mathbf{a} + Etc.$$

$$= \mathbf{a}^T \mathbf{C}_{\mathbf{xx}}\mathbf{a} + Etc.$$

$$= \underbrace{\mathbf{a}^T \mathbf{C}_{\mathbf{xx}}\mathbf{a}}_{\substack{\text{covariance} \\ \textit{matrix}}} \underbrace{- \mathbf{a}^T \mathbf{c}_{\mathbf{x}\theta} - \mathbf{c}_{\theta\mathbf{x}}\mathbf{a}}_{\substack{\text{cross-covariance} \\ \textit{vectors}}} + \underbrace{c_{\theta\theta}}_{\text{variance}}$$

$N\times N$ $\quad N\times 1$ $\quad 1\times N$ $\quad 1\times 1$

$$\boxed{\begin{array}{c} \mathbf{c}_{\mathbf{x}\theta} = E\left\{\left(\mathbf{x} - E\{\mathbf{x}\}\right)\left(\theta - E\{\theta\}\right)\right\} \qquad \mathbf{c}_{\theta\mathbf{x}} = E\{\left(\theta - E\{\theta\}\right)\left(\mathbf{x} - E\{\mathbf{x}\}\right)^T\} \\ c_{\theta\theta} = E\left\{\left(\theta - E\{\theta\}\right)^2\right\} \end{array}}$$

Note: $\mathbf{c}_{\theta\mathbf{x}}^T = \mathbf{c}_{\mathbf{x}\theta}$ $\Rightarrow$ $\boxed{Bmse(\hat{\theta}) = \mathbf{a}^T \mathbf{C}_{\mathbf{xx}}\mathbf{a} - 2\mathbf{a}^T \mathbf{c}_{\mathbf{x}\theta} + c_{\theta\theta}}$

Step #3: Minimize w.r.t. $a_1, a_2, \ldots, a_{N\text{-}1}$

Only up to N-1

$$\frac{\partial Bmse(\hat{\theta})}{\partial \mathbf{a}} = 0$$

$$2\mathbf{C}_{\mathbf{xx}}\mathbf{a} - 2\mathbf{c}_{\mathbf{x}\theta} = \mathbf{0} \quad \Rightarrow \quad \boxed{\mathbf{a} = \mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{c}_{\mathbf{x}\theta}} \qquad \mathbf{a}^T = \mathbf{c}_{\theta\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}$$

This is where the statistical dependence between the data and the parameter is used… via a cross-covariance vector

Step #4: Combine Results

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n] + a_N$$

$$= \mathbf{a}^T\mathbf{x} + \left[E\{\theta\} - \mathbf{a}^T E\{\mathbf{x}\}\right] = E\{\theta\} + \mathbf{a}^T\left(\mathbf{x} - E\{\mathbf{x}\}\right)$$

So the Optimal LMMSE Estimate is:

$$\boxed{\hat{\theta} = E\{\theta\} + \mathbf{c}_{\theta\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}\left(\mathbf{x} - E\{\mathbf{x}\}\right)}$$

If Means = 0

$$\boxed{\hat{\theta} = \mathbf{c}_{\theta\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{x}}$$

Note: LMMSE Estimate Only Needs 1st and 2nd Moments… not PDFs!!

Step #5: Find Minimum Bmse

Substitute into Bmse result and simplify:

$$Bmse(\hat{\theta}) = \mathbf{a}^T \mathbf{C}_{\mathbf{xx}} \mathbf{a} - 2\mathbf{a}^T \mathbf{c}_{\mathbf{x}\theta} + c_{\theta\theta}$$

$$= \mathbf{c}_{\theta\mathbf{x}} \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{C}_{\mathbf{xx}} \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{c}_{\mathbf{x}\theta} - 2\mathbf{c}_{\theta\mathbf{x}} \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{c}_{\mathbf{x}\theta} + c_{\theta\theta}$$

$$= \mathbf{c}_{\theta\mathbf{x}} \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{c}_{\mathbf{x}\theta} - 2\mathbf{c}_{\theta\mathbf{x}} \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{c}_{\mathbf{x}\theta} + c_{\theta\theta}$$

$$\boxed{Bmse(\hat{\theta}) = c_{\theta\theta} - \mathbf{c}_{\theta\mathbf{x}} \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{c}_{\mathbf{x}\theta}}$$

Note: If θ and **x** are statistically independent then $\mathbf{C}_{\theta\mathbf{x}} = \mathbf{0}$

$$\boxed{\hat{\theta} = E\{\theta\}}$$

$$\boxed{Bmse(\hat{\theta}) = c_{\theta\theta}}$$

Totally based on prior info… the data is useless

Ex. 12.1 DC Level in WGN with Uniform Prior

Recall: Uniform prior gave a non-closed form requiring integration

…but changing to a Gaussian prior fixed this.

Here we keep the uniform prior and get a simple form:

- by using the Linear MMSE

For this problem the LMMSE estimate is: $\hat{A} = \mathbf{c}_{A\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{x}$

Need

$$\mathbf{C}_{\mathbf{xx}} = E\left\{(A\mathbf{1}+\mathbf{w})(A\mathbf{1}+\mathbf{w})^T\right\} = \sigma_A^2\mathbf{1}\mathbf{1}^T + \sigma^2\mathbf{I}$$

$$\mathbf{c}_{\theta\mathbf{x}} = E\{A\mathbf{x}\} = E\left\{A(A\mathbf{1}+\mathbf{w})^T\right\} = \sigma_A^2\mathbf{1}^T$$

A & $\mathbf{w}$ are uncorrelated

$$\boxed{\hat{A} = \left[\frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N}\right]\bar{x}}$$

12.4 Geometrical Interpretations

Abstract Vector Space

Mathematicians first tackled "physical" vector spaces like R^N and C^N, etc.

But… then abstracted the "bare essence" of these structures into the general idea of a vector space.

We've seen that we can interpret Linear LS in terms of "Physical" vector spaces.

We'll now see that we can interpret Linear MMSE in terms of "Abstract" vector space ideas.

Abstract Vector Space Rules

An abstract vector space consists of a set of “mathematical objects” called vectors and another set called scalars that obey:

1. There is a well-defined operation of “addition” of vectors that gives a vector in the set, and…
   - “Adding” is commutative and associative
   - There is a vector in the set – call it **0** – for which “adding” it to any vector in the set gives back that same vector
   - For every vector there is another vector s.t. when the 2 are added you get the **0** vector
2. There is a well-defined operation of “multiplying” a vector by a “scalar” and it gives a vector in the set, and…
   - “Multiplying” is associative
   - Multiplying a vector by the scalar 1 gives back the same vector
3. The distributive property holds
   - Multiplication distributes over vector addition
   - Multiplication distributes over scalar addition

Examples of Abstract Vector Spaces

1. Scalars = Real Numbers
   Vectors = N^{th} Degree Polynomials w/ Real Coefficients

2. Scalars = Real Numbers
   Vectors = $M \times N$ Matrices of Real Numbers

3. Scalars = Real Numbers
   Vectors = Functions from [0,1] to R

4. Scalars = Real Numbers
   Vectors = Real-Valued Random Variables with Zero Mean

Colliding Terminology…
a *scalar* RV is a *vector*!!!

Inner Product Spaces

An extension of the idea of Vector Space… must also have:

There is a well-defined concept of inner product s.t. all the rules of "ordinary" inner product still hold

Not needed for Real IP Spaces

- $<\mathbf{x},\mathbf{y}> \; = \; <\mathbf{y}, \mathbf{x}>^*$
- $<a_1\mathbf{x}_1 + a_2\mathbf{x}_2,\mathbf{y}> \; = \; a_1<\mathbf{x}_1,\mathbf{y}> \; + \; a_2<\mathbf{x}_2,\mathbf{y}>$
- $<\mathbf{x},\mathbf{x}> \; \geq \; 0; <\mathbf{x},\mathbf{x}> \; = \; 0 \;\; \text{iff} \;\; \mathbf{x} = \mathbf{0}$

Note: an inner product "induces" a norm (or length measure):

$$||\mathbf{x}||^2 = <\mathbf{x},\mathbf{x}>$$

So an inner product space has:

1. Two sets of elements: Vectors and Scalars
2. Algebraic Structure (Vector Addition & Scalar Multiplication)
3. Geometric Structure
   - Direction (Inner Product)
   - Distance (Norm)

Inner Product Space of Random Variables

Vectors: Set of all real RVs w/ zero mean & finite variance (ZMFV)

Scalars: Set of all real numbers

Inner Product: $<X,Y> = E\{XY\}$

Inner Product is Correlation!
Uncorrelated = Orthogonal

Claim… This is an Inner Product Space

First… this is a vector space…

Addition Properties: $X+Y$ is another ZMFV RV

1. It is Associative and Commutative: $X+(Y+Z) = (X+Y)+Z;\ X+Y = Y+X$
2. The zero RV has variance of 0 (What is an RV with var = 0???)
3. The negative of RV X is $-X$

Multiplication Properties: For any real # a, aX is another ZMFV RV

1. It is Associative: $a(bX) = (ab)X$
2. $1X = X$

Distributive Properties:

1. $a(X+Y) = aX + aY$
2. $(a+b)X = aX + bX$

Next…This is an inner product space…

- $<a_1X_1 + a_2X_2, Y> = E\{(a_1X_1 + a_2X_2)Y\}$
  $= a_1E\{X_1Y\} + a_2E\{X_2Y\}$
- $||X||^2 = <X, X> = E\{X^2\} = \text{var}\{X\} \geq 0$

Use IP Space Ideas for Section 12.3

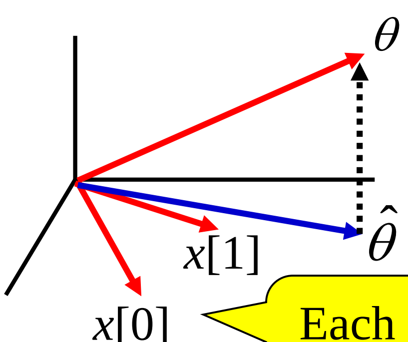
Apply to the Estimation of a zero-mean scalar RV: $\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n]$

Zero-Mean… don't need a_N

Trying to estimate the realization of RV θ via a linear combination of N other RVs $x[0]$, $x[1]$, $x[2]$,… $x[N\text{-}1]$

Now…using our new vector space view of RVs, this is the same *structural* mathematics that we saw for the Linear LS !

N = 2 Case



Each RV is *viewed* as a vector

Minimize: $\underbrace{\left\| \theta - \hat{\theta} \right\|^2}_{\text{Connects to Geometry}} = \underbrace{E\left\{ \left(\theta - \hat{\theta} \right)^2 \right\} = Bmse\left(\hat{\theta} \right)}_{\text{Connects to MSE}}$

Recall Orthogonality Principle!!!

Estimation Error ⊥ Data Space

$$\boxed{E\left\{ \left(\theta - \hat{\theta} \right) x[n] \right\} = 0}$$

Now apply this Orthogonality Principle…

$$E\left\{(\theta-\hat{\theta})\mathbf{x}^T\right\}=\mathbf{0}^T \quad \text{with} \quad \hat{\theta}=\mathbf{a}^T\mathbf{x}$$

$$E\left\{(\theta-\mathbf{a}^T\mathbf{x})\mathbf{x}^T\right\}=\mathbf{0}^T \quad \Rightarrow \quad E\left\{\theta\mathbf{x}^T\right\}=\mathbf{a}^T E\left\{\mathbf{x}\mathbf{x}^T\right\} \quad \Rightarrow \quad E\left\{\mathbf{x}\theta^T\right\}=E\left\{\mathbf{x}\mathbf{x}^T\right\}\mathbf{a}$$

$$\boxed{\mathbf{C}_{\mathbf{xx}}\mathbf{a}=\mathbf{c}_{\mathbf{x}\theta}}$$ "*The* Normal Equations"

Assuming that $\mathbf{C}_{\mathbf{xx}}$ is invertible…

$$\boxed{\mathbf{a}=\mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{c}_{\mathbf{x}\theta}} \Rightarrow \boxed{\hat{\theta}=\mathbf{a}^T\mathbf{x}=\mathbf{c}_{\theta\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{x}}$$

Same as before!!!

12.5 Vector LMMSE Estimator

Meaning a "Physical" Vector

Estimate: Realization of $\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_p \end{bmatrix}^T$

Linear Estimator: $\hat{\boldsymbol{\theta}} = \mathbf{A}\mathbf{x} + \mathbf{a}$

Goal: Minimize Bmse ***for each element***

⇒ View i^{th} row in **A** and i^{th} element in ***a*** as forming a scalar LMMSE estimator for θ_i

Already know the individual element solutions!

- **Write them down**
- **Combine into matrix form**

Solutions to Vector LMMSE

The Vector LMMSE estimate is:

$$\boxed{\hat{\boldsymbol{\theta}} = E\{\boldsymbol{\theta}\} + \mathbf{C}_{\boldsymbol{\theta}\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}[\mathbf{x} - E\{\mathbf{x}\}]}$$

Now… $p \times N$ Matrix… Cross-Covariance Matrix

Still… $N \times N$ Matrix… Covariance Matrix

If $E\{\boldsymbol{\theta}\} = \mathbf{0}$ & $E\{\mathbf{x}\} = \mathbf{0}$ $\Rightarrow$ $\boxed{\hat{\boldsymbol{\theta}} = \mathbf{C}_{\boldsymbol{\theta}\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{x}}$

Can show similarly that Bmse <u>Matrix</u> is

$$\mathbf{M}_{\hat{\boldsymbol{\theta}}} = E\{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T\}$$

$$\boxed{\mathbf{M}_{\hat{\boldsymbol{\theta}}} = \mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}} - \mathbf{C}_{\boldsymbol{\theta}\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{C}_{\mathbf{x}\boldsymbol{\theta}}}$$

$p \times p$
prior Cov. Matrix

$p \times N$

$N \times N$

$N \times p$

Two Properties of LMMSE Estimator

1. Commutes over affine transformations

If $\boldsymbol{\alpha} = \mathbf{A}\boldsymbol{\theta} + \mathbf{b}$ and $\hat{\boldsymbol{\theta}}$ is LMMSE Estimate

Then $\hat{\boldsymbol{\alpha}} = \mathbf{A}\hat{\boldsymbol{\theta}} + \mathbf{b}$ is LMMSE Estimate for $\boldsymbol{\alpha}$

2. If $\boldsymbol{\alpha} = \boldsymbol{\theta}_1 + \boldsymbol{\theta}_2$ then $\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\theta}}_1 + \hat{\boldsymbol{\theta}}_2$

Bayesian Gauss-Markov Theorem

Like G-M Theorem for the BLUE

Let the data be modeled as $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$

known

$p\times1$ random
mean $\boldsymbol{\mu}_\theta$
Cov Mat $\mathbf{C}_{\theta\theta}$
(Not Gaussian)

$N\times1$ random
zero mean
Cov Mat $\mathbf{C}_w$
(Not Gaussian)

Application of previous results, evaluated for this data model gives:

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\mu}_\theta + \mathbf{C}_{\theta\theta}\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_{\theta\theta}\mathbf{H}^T + \mathbf{C}_w\right)^{-1}[\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_\theta]$$

$$\mathbf{C}_\varepsilon = \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta\theta}\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_{\theta\theta}\mathbf{H}^T + \mathbf{C}_w\right)^{-1}\mathbf{H}\mathbf{C}_{\theta\theta}$$

MMSE Matrix:

$$\mathbf{M}_{\hat{\theta}} = \mathbf{C}_\varepsilon$$

Same forms as for Bayesian Linear Model (which include Gaussian assumption)

Except here… the result is suboptimal… unless the optimal estimate *is* linear

In practice… generally don't know if linear estimate is optimal… but we use LMMSE for its simple form!

The challenge is to "guess" or estimate the needed means & cov matrices