Home exam in Randomized Algorithms
(with solutions)

2004.05.24-2004.06.03

---

1. Markov chain theory

Let $(X_0, X_1, ...)$ be a Markov chain with state space $S = \{s_1, ..., s_k\}$ and transition matrix P. Suppose P is doubly stochastic. What can we say about the asymptotic distribution of the chain? Determine if the chain ...

(a) ... has a stationary distribution, and if it does, specify it.

(b) ... is irreducible.

(c) ... is aperiodic.

(d) ... has convergence towards a stationary distribution.

For each property, a proof or counterexample is required.

Solution:

(a) Let P be a double stochastic matrix and let λ be the uniform distribution on S.

$$\lambda = \begin{pmatrix} \frac{1}{k} & \frac{1}{k} & \cdots & \frac{1}{k} \end{pmatrix}$$

Then λ is a stationary distribution for P if $\lambda P = \lambda$. We perform the calculation to get the following ...

$$\begin{aligned}
\lambda A &= \begin{pmatrix} \frac{1}{k} & \frac{1}{k} & \cdots & \frac{1}{k} \end{pmatrix} \begin{pmatrix} p_{1,1} & p_{1,2} & \cdot & \cdot & \cdot & p_{1,k} \\ p_{2,1} & p_{2,2} & \cdot & \cdot & \cdot & p_{2,k} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ p_{k,1} & p_{k,2} & \cdot & \cdot & \cdot & p_{k,k} \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^{k} \frac{p_{i,1}}{k} & \sum_{i=1}^{k} \frac{p_{i,2}}{k} & \cdots & \sum_{i=1}^{k} \frac{p_{i,k}}{k} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{k} \sum_{i=1}^{k} p_{i,1} & \frac{1}{k} \sum_{i=1}^{k} p_{i,2} & \cdots & \frac{1}{k} \sum_{i=1}^{k} p_{i,k} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{k} & \frac{1}{k} & \cdots & \frac{1}{k} \end{pmatrix} = \lambda
\end{aligned}$$

... and we are done.

(b) Counterexample: The following matrix

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

is the transition matrix of a reducible chain, since for all $n \geq 0$ we have $(P^n)_{1,2} = P_{1,2} = 0$.

(c) Counterexample: The following matrix

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is the transition matrix of a periodic chain, since for $k \geq 0$

$$P^{2k+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \;, \quad P^{2k} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

(d) In general we will not have convergence towards a statioanry distribution for double stochastic matrices. Again the matrix

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is a counterexample. Stared in distribution $\lambda = (\lambda_1 \;\; \lambda_2)$ distribution

$$\mu^{(n)} = (\lambda_1 \;\; \lambda_2)$$

for $n \in \{0, 2, 4, ...\}$ and

$$\mu^{(n)} = (\lambda_2 \;\; \lambda_1)$$

for $n \in \{1, 3, 5, ...\}$.

2. General Markov Chain Monte Carlo

Let V and S be two finite sets and let π be a probability distribution on S^V. We call an element $\xi \in S^V$ a configuration. For any $w \in V$ we let $\xi(w)$ denote the configuration at w and we let $\xi^w \in S^{V \setminus \{w\}}$ denote the configuration on every element in V except w.

When constructing a Gibbs sampler to generate samples from π we update a configuration ξ to ξ' according to the following procedure.

1. Pick an element $v \in V$ uniformly at random.
2. Let $\xi'(v) = a$ with probability $\pi(\xi'(v) = a|\xi^v)$ for any $a \in S$.
3. For $w \neq v$ let $\xi'(w) = \xi(w)$

Show that this in general makes π a reversible distribution.

Solution: For any two adjacent states ξ, ξ' in the Markov chain we must show that

$$\pi(\xi'(v) = b \ |\xi^v)\pi(\xi) = \pi(\xi(v) = a \ |(\xi')^v)\pi(\xi')$$

Let ξ, ξ' be two states in S^V differing in one vertex $v \in V$ only. Let $\xi(v) = a$ and $\xi'(v) = b$ and $\xi(w) = \xi'(w)$ whenever $w \neq v$.

$$\begin{aligned}
\pi(\xi'(v) = b \ |\xi^v)\pi(\xi) &= \frac{\pi(\xi'(v) = b, \ \xi^v)}{\pi(\xi^v)} \ \pi(\xi) \\
&= \frac{\pi(\xi'(v) = b, \ \xi^v)}{\pi(\xi^v)} \ \pi(\xi(v) = a, \xi^v) \\
&= \frac{\pi(\xi(v) = a, \ \xi^v)}{\pi(\xi^v)} \ \pi(\xi'(v) = b, \xi^v) \\
&\overset{\xi^v = (\xi')^v}{=} \frac{\pi(\xi(v) = a, \ (\xi')^v)}{\pi((\xi')^v)} \ \pi(\xi'(v) = b, (\xi')^v) \\
&= \pi(\xi(v) = a \ |(\xi')^v)\pi(\xi')
\end{aligned}$$

3. An MCMC Application

A simple version of the so called Widom-Rowlinson model is defined as follows. Let G be a finite graph (for instance, a finite portion of the square lattice), with vertex set V. An assignment $\xi \in \{-1, 0, +1\}^V$ of -1's, 0's, and +1's to the vertices of G, is said to be WR-feasible if $\xi(u)\xi(v) \geq 0$ for all pairs of vertices that share an edge. In other words, ξ is WR-feasible if no edge has a -1 at one end point and a +1 at the other. The Widom-Rowlinson model now arises by picking a WR-feasible assignment at random, with equal probability for each such assignment. Write μ_G for the corresponding probability distribution on $\{-1, 0, +1\}^V$.

Propose a sensible MCMC algorithm for simulating μ_G, and verify that it has the desired convergence property.

Solution: We use the Gibbs sampler. Given X_n we update to X_{n+1} according to the following scheme.

1. Pick a vertex $v \in V$ uniformly at random.
2. Let $X_{n+1}(v) = a$ accrding to the conditional distribution $\mathbb{P}(X_{n+1}(v) = a \mid X_n^v)$ for $a \in \{-1, 0 - +1\}$.
3. For $w \in V \setminus \{v\}$ let $X_{n+1}(w) = X_n(w)$.

Remain to compute the conditional distribution in step 2. Let μ_G be the probability distribution according to the Widom-Rowlinson model described above. This is the uniform distribution over the set of WR-feasible configurations in $\Omega = \{-1, 0, +1\}^V$. We can write this as

$$\mu_G(\xi) = \frac{1}{Z} \prod_{\langle x,y \rangle \in E} I_{\xi(x)\xi(y) \geq 0}(x, y)$$

for any $\xi \in \Omega$, where Z is a normalising constant. Given $v \in V$ let

$$V_v = \{w \in V : \langle w, v \rangle \in E\}$$

be the set of vertices adjacent to v, also let E^v consist of edges not having v as an end vertex.

$$\begin{aligned}
\mu_G(X_{n+1}(v) = +1 \mid X_n^v) &= \frac{\mu_G(X_{n+1}(v) = +1,\ X_n^v)}{\sum_{c \in \{-1,0,+1\}} \mu_G(X_n(v) = c,\ X_n^v)} \\
&= \left(\sum_{c \in \{-1,0,+1\}} \frac{\mu_G(X_n(v) = c,\ X_n^v)}{\mu_G(X_{n+1}(v) = +1,\ X_n^v)} \right)^{-1} \\
&= \left(\frac{\mu_G(X_n(v) = -1,\ X_n^v)}{\mu_G(X_{n+1}(v) = +1,\ X_n^v)} + \frac{\mu_G(X_n(v) = 0,\ X_n^v)}{\mu_G(X_{n+1}(v) = +1,\ X_n^v)} + 1 \right)^{-1}
\end{aligned}$$

Now consider one of the terms in the paranthesis.

$$
\begin{aligned}
\frac{\mu_G(X_n(v)=-1,\ X_n^v)}{\mu_G(X_{n+1}(v)=+1,\ X_n^v)} &= \frac{\prod_{w\in V_v} I_{\{X_n(w)X_n(v)\geq 0\}}(w,v)\ \prod_{\langle x,y\rangle\in E^v} I_{\{X_n(x)X_n(y)\geq 0\}}(x,y)}{\prod_{w\in V_v} I_{\{X_{n+1}(w)X_{n+1}(v)\geq 0\}}(w,v)\ \prod_{\langle x,y\rangle\in E^v} I_{\{X_n(x)X_n(y)\geq 0\}}(x,y)} \\
&= \frac{\prod_{w\in V_v} I_{\{X_n(w)X_n(v)\geq 0\}}(w,v)}{\prod_{w\in V_v} I_{\{X_{n+1}(w)X_{n+1}(v)\geq 0\}}(w,v)} \\
&= \frac{\prod_{w\in V_v} I_{\{-X_n(w)\geq 0\}}(w,v)}{\prod_{w\in V_v} I_{\{X_{n+1}(w)\geq 0\}}(w,v)}
\end{aligned}
$$

If we put everything together we get the following

$$
\mu_G(X_{n+1}(v)=+1 \mid X_n^v) = \left(\frac{\prod_{w\in V_v} I_{\{X_n(w)\leq 0\}}(w,v)}{\prod_{w\in V_v} I_{\{X_{n+1}(w)\geq 0\}}(w,v)} + \frac{1}{\prod_{w\in V_v} I_{\{X_{n+1}(w)\geq 0\}}(w,v)} + 1\right)^{-1}
$$

Now consider a neighbourhood containing only 0's, then we get the following conditional probabilities.

$$\mu_G(X_{n+1}(v)=-1 \mid X_n^v) = \frac{1}{3}$$

$$\mu_G(X_{n+1}(v)=0 \mid X_n^v) = \frac{1}{3}$$

$$\mu_G(X_{n+1}(v)=+1 \mid X_n^v) = \frac{1}{3}$$

For a neighbourhood containing -1 and 0 we get.

$$\mu_G(X_{n+1}(v)=-1 \mid X_n^v) = \frac{1}{2}$$

$$\mu_G(X_{n+1}(v)=0 \mid X_n^v) = \frac{1}{2}$$

$$\mu_G(X_{n+1}(v)=+1 \mid X_n^v) = 0$$

For a neighbourhood containing $+1$ and 0 we get.

$$\mu_G(X_{n+1}(v)=-1 \mid X_n^v) = 0$$

$$\mu_G(X_{n+1}(v)=0 \mid X_n^v) = \frac{1}{2}$$

$$\mu_G(X_{n+1}(v)=+1 \mid X_n^v) = \frac{1}{2}$$

For a neighbourhood containing at least one -1 and one $+1$ we get.

$$\mu_G(X_{n+1}(v) = -1 \mid X_n^v) = 0$$

$$\mu_G(X_{n+1}(v) = 0 \mid X_n^v) = 1$$

$$\mu_G(X_{n+1}(v) = +1 \mid X_n^v) = 0$$

So the conditional distribution of the configuration at a certain site given the rest is the uniform distribution over the set of values in $\{-1, 0, +1\}$ feasible at v.

We now get the following algorithm. Let X_0 be an arbitrary state in $\{-1, 0, +1\}^V$, for example the one assigning a 0 to every vertex in V. Given $X_n,\ n \geq 0$ we update to X_{n+1} by the following mechanism.

1. Pick a vertex $v \in V$ uniformly at random.
2. Assign $X_{n+1}(v)$ a value according to the uniform distribution over the set of feasible values in $\{-1, 0, +1\}$.
3. For $w \in V \setminus \{v\}$ let $X_{n+1}(w) = X_n(w)$.

This is the Gibbs sampler for the Widom-Rowlinson model and it has μ_G as stationary distribution according to assignment 2. The chain is irreducible and apriodic.

Irreducibility: Let $\xi \in \{-1, 0, +1\}^V$ be an arbitray state. Let $A = \{v_1, ..v_k\}$ be the set of vertices in V having a configuration other than 0. The probability of choosing the sequence $v_1, ..., v_k$ for our next k updates is $(1/|V|)^k > 0$. At every such update we have probability of at least $1/3$ to assign vertex v_i a 0. The probability of moving from ξ to the "all zero" state is at least

$$\left(\frac{1}{3|V|}\right)^k > 0.$$

We can use the same argument to show that we have a positive probability of moving from the "all zer0" state to any other state in $k' < \infty$ steps. Thus the chain is irreducible.

Apriodicity: Suppose X_n is the "all zero" state, the with probability $1/3$ x_{n+1} is also the "all zero" state. This makes the "all zero" state apriodic, and aperiodicity for all other states follows by irreducibility.

Now the wanted convergence property follows from theorem 5.2 and we are done.

4. Sampling of Markov chains

Let $(X_0, X_1, ...)$ be a Markov chain with transition matrix P.

(a) Define $(Y_0, Y_1, ...)$ by setting $Y_n = X_{2n}$ for each n. Show that $(Y_0, Y_1, ...)$ is a Markov chain with transition matrix P^2.

(b) Find an appropriate generalization of the result in (a) to the situation where we sample every k^{th} (rather than every second) value of $(X_0, X_1, ...)$.

Solution:

(a) First we show the Markov property for the $(Y_0, Y_1, ...)$ chain, this is a consequence of the Markov property of the $(X_0, X_1, ...)$ chain. Fix an arbitrary $n \geq 1$.

$$\begin{aligned}
&\mathbb{P}(Y_{n+1} = s_{n+1} | Y_0 = s_0,\ Y_1 = s_1, ...,\ Y_n = s_n)\\
&= \mathbb{P}(X_{n+2} = s_{n+2} | X_0 = s_0,\ X_2 = s_2,\ X_4 = s_4, ...,\ X_{2n} = s_{2n})\\
&\overset{(1)}{=} \mathbb{P}(X_{n+2} = s_{n+2} | X_{2n} = s_{2n})\\
&= \mathbb{P}(Y_{n+1} = s_{n+1} | Y_n = s_n)
\end{aligned}$$

where (1) follows since $(X_0, X_1, ...)$ is a Markov chain. This makes $(Y_0, Y_1, ...)$ a Markov chain.

Remains to show that the transition probability is the required one. Consider the probability of a transition in two steps bewteen states s_i and s_j, $\mathbb{P}(X_{n+2} = s_j | X_n = s_i)$. We condition on X_{n+1} and use the law of total probability.

$$\begin{aligned}
\mathbb{P}(X_{n+2} = s_j | X_n = s_i) &= \sum_{s_l \in S} \mathbb{P}(X_{n+2} = s_j, X_{n+1} = s_l | X_n = s_i)\\
&= \sum_{s_l \in S} \underbrace{\mathbb{P}(X_{n+1} = s_l | X_n = s_i)}_{=P_{i,l}} \underbrace{\mathbb{P}(X_{n+2} = s_j | X_{n+1} = s_l)}_{=P_{l,j}}\\
&= \sum_{s_l \in S} P_{i,l} P_{l,j}\\
&= P^2_{i,j}
\end{aligned}$$

(b) Let $k \geq 1$. We start with the Markov propery. Fix an arbitrary $n \geq 1$.

$$\begin{aligned}
&\mathbb{P}(Y_{n+1} = s_{n+1} | Y_0 = s_0,\ Y_1 = s_1, ...,\ Y_n = s_n)\\
&= \mathbb{P}(X_{n+k} = s_{n+k} | X_0 = s_0,\ X_k = s_k,\ X_{2k} = s_{2k}, ...,\ X_{nk} = s_{nk})\\
&\overset{(1)}{=} \mathbb{P}(X_{n+k} = s_{n+k} | X_{nk} = s_{nk})\\
&= \mathbb{P}(Y_{n+1} = s_{n+1} | Y_n = s_n)
\end{aligned}$$

where (1) (again) follows since $(X_0, X_1, ...)$ is a Markov chain. This makes $(Y_0, Y_1, ...)$ a Markov chain.

Remains to show that $(Y_0, Y_1, ...)$ has the wanted transition matrix. We need to show $\mathbb{P}(X_{n+k} = s_j | X_n = s_i) = P^k_{i,j}$. We have already done the base case in (a). Now assum that the statement is true for $k = m$ and consider transition probability $\mathbb{P}(X_{n+m+1} = s_j | X_n = s_i)$.

$$\begin{aligned}
\mathbb{P}(X_{n+m+1} = s_j | X_n = s_i) &= \sum_{s_l \in S} \mathbb{P}(X_{n+2} = s_j, X_{n+m} = s_l | X_n = s_i) \\
&= \sum_{s_k \in S} \underbrace{\mathbb{P}(X_{n+m} = s_l | X_n = s_i)}_{=P^m_{i,l}} \underbrace{\mathbb{P}(X_{n+2} = s_j | X_{n+m} = s_l)}_{=P_{l,j}} \\
&= \sum_{s_l \in S} P^m_{i,l} P_{l,j} \\
&= P^{m+1}_{i,j}
\end{aligned}$$

And we are done.

5. Convergence of Markov chains

Let $(X_0, X_1, ...)$ be an irreducible and aperiodic Markov chain with state space $S = \{s_1, ..., s_k\}$ and transition matrix P. In general P has k eigenvalues and k left eigenvectors. Let $\lambda_1, \lambda_2, ..., \lambda_k$ denote the eigenvalues in increasing order. If P is stochastic the largest eigenvalue $\lambda_k = 1$.

(a) What can you say about the rows of P^n as $n \to \infty$?

(b) Which is the corresponding left eigenvector to λ_k?

(c) What can you say about the convergence of $\mu^{(n)}$ towards the stationary distribution? Do the eigenvalues have any implication for the convergence rate?

Solution:

(a) Since the chain is is irreducible and aperiodic there exists a stationary distribution. Let π be the stationary distribution and let

$$Q = \lim_{n\to\infty} P^n.$$

We know that regardless of initial distribution $\mu^{(0)}$ we have $\mu^{(0)}Q = \pi$. Let δ_i be the probability distribibution that put all probability on state s_i, for any $i \in \{1, ..., k\}$.

$$\delta_i(s) = \begin{cases} 1, & s = s_i \\ 0, & \text{otherwise} \end{cases}$$

Let Q_i be the i:th row in Q. For any $i \in \{1, ..., k\}$ we have

$$\delta_i Q = Q_i = \pi.$$

As $n \to \infty$ the rows of Q converges to π.

(b) Let $\lambda_1, ..., \lambda_k$ be the eigenvalues in increasing order of the transition matrix P. The largest eigenvalue of a stochastic eigenvector is 1 so we have $\lambda_k = 1$. Let v_k^l be the left eigenvector corresponding to λ_k then

$$v_k^l P = v_k^l$$

and we can deduce that v_k^l is the stationary distribution of the chain having P as transition matrix.

(c) As $n \to \infty$ we have $\mu^{(0)} P^n \to \pi$. We can rewrite P^n as VD^nV^{-1}.

$$
\begin{aligned}
VD^nV^{-1} &= V \begin{pmatrix} \lambda_1^n & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & \lambda_2^n & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \lambda_{k-1}^n & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 1 \end{pmatrix} V^{-1} \\
&= V \begin{pmatrix} 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 1 \end{pmatrix} V^{-1} + V \begin{pmatrix} \lambda_1^n & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & \lambda_2^n & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \lambda_{k-1}^n & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \end{pmatrix} V^{-1}
\end{aligned}
$$

since $\lambda_k = 1$, and $\lambda_1, ..., \lambda_{k-1}$ all are stricty smaller than 1, we have

$$
\begin{pmatrix} \lambda_1^n & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & \lambda_2^n & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \lambda_{k-1}^n & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \end{pmatrix} \xrightarrow{n \to \infty} \begin{pmatrix} 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \end{pmatrix}
$$

For all $i \in \{1, ..., k\}$ $\lambda_i^n \to 0$ as $n \to \infty$. The smaller λ_i is the faster the convergence towards 0 is. So the rate of convergence of towards the stationary distribution is determined bythe second largest eigenvalue, λ_{k-1}.

SOME RESULTS AND DEFINITIONS FROM LINEAR ALGEBRA

Stochastic and doubly stochastic matrices :

A $n \times n$ real matrix M is called stochastic if

$$\forall i \in \{1, ..., n\} : \sum_{j=1}^{n} M_{i,j} = 1.$$

If in addition

$$\forall j \in \{1, ..., n\} : \sum_{i=1}^{n} M_{i,j} = 1.$$

holds we say that M is doubly stochastic. A doubly stochastic matrix is a matrix where all rows and all columns sum up to 1.

Eigenvalues and eigenvectors of stochastic matrices :

A $n \times n$ real matrix M has in general n eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ and to each eigenvalue there is a corresponding right eigenvector v_k^r and left eigenvector v_k^l and in general $v_k^r \neq v_k^l, \ k = 1, ..., n$.

We can factorise M as follows.

$$M = V \ D \ V^{-1}$$

Where D is a diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \lambda_2 & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \lambda_n \end{pmatrix}$$

and V is a matrix having the eigenvectors as columns

$$V = \begin{pmatrix} v_1^r & v_2^r & v_3^r & \cdot & \cdot & \cdot & v_n^r \end{pmatrix}$$