

# Math behind Country-Capability-Product Decomposition

ANDRES GOMEZ-LIEVANO

Center for International Development, HKS, Harvard University

## 1 Statement of the Problem

Our starting point is the empirical matrix  $\mathbf{M}$  that identifies the products that countries export with a relative comparative advantage:

$$[\mathbf{M}]_{c,p} = 1, \quad \text{if } RCA_{c,p} > 1.$$

The empirical matrix we observe whereby some countries are able to produce some products we hypothesize comes from an interaction between a matrix  $\mathbf{C}_{ca}$  of countries and different capabilities they may have, and another matrix  $\mathbf{P}_{ap}$  specifying the capabilities that products require to get produced. The problem is to infer these two matrices from  $\mathbf{M}$  plus a minimum number of assumptions.

## 2 Framework: Stochastic Matrices

Before we tackle directly our problem, let us show that we can manipulate more easily our matrices if we represent them as stochastic matrices. The reason for this is that the product of stochastic matrices is still a stochastic matrix. And stochastic matrices work, from a certain point of view, as matrices that translate a state of the world into another.

Let us consider the bipartite network representing  $\mathbf{M}$ , which connects countries with products. And let us consider a random walker in this network.

Suppose there is a probability  $\pi_p(t)$  that the random walker is in product  $p$  at time  $t$ . The probability that the random walker will end up at country  $c$  at  $t + 1$  (after randomly following one of the edges on the bipartite network) can be expressed as:

$$\begin{aligned} & \Pr\{ \text{ending at } c \text{ at } t + 1 \} \\ &= \sum_p \Pr\{ \text{ending at } c \text{ at } t + 1 \mid \text{starting at } p \text{ at } t \} \Pr\{ \text{starting at } p \text{ at } t \}. \end{aligned}$$

In terms of our matrix  $\mathbf{M}$ , and given the vector  $\vec{\pi}(t)$  of starting in any of the products, the previous equation becomes

$$\gamma_c(t + 1) = \sum_p \left( \frac{M_{c,p}}{\sum_{c'} M_{c',p}} \right) \pi_p(t), \quad (1)$$

where  $\vec{\gamma}(t + 1)$  is the vector of probabilities of being in the different countries at the time-step  $t + 1$ . In matrix notation, we write this as

$$\vec{\gamma}(t + 1) = \mathbf{Q}_{c \leftarrow p} \cdot \vec{\pi}(t). \quad (2)$$

I have explicitly specified that the stochastic matrix  $\mathbf{Q}$  is a transition matrix from products to countries (i.e., “ $c \leftarrow p$ ”), and is thus a *left-stochastic* matrix<sup>1</sup>.

We can do a similar exercise but starting from countries and ending up at products. This yields

$$\vec{\pi}^T(t + 1) = \vec{\gamma}^T(t) \cdot \mathbf{R}_{c \rightarrow p}, \quad (3)$$

where  $(\cdot)^T$  is the transpose operation,  $\mathbf{R}$  is the transition matrix that takes me from countries to products (i.e., “ $c \rightarrow p$ ”), and is thus a *right-stochastic* matrix:

$$\mathbf{R}_{c \rightarrow p} = \left( \frac{M_{c,p}}{\sum_{p'} M_{c,p'}} \right).$$

---

<sup>1</sup>A left-stochastic matrix has the property that its columns add up to 1. We transition from states and their probabilities to future states by multiplying on the right by a vector of probabilities for the previous states.

Transition matrices  $\mathbf{Q}_{c \leftarrow p}$  and  $\mathbf{R}_{c \rightarrow p}$  can be expressed explicitly in terms of  $\mathbf{M}$  as

$$\mathbf{Q} = \mathbf{M} \cdot \text{diag}(\vec{\mathbf{1}}^T \cdot \mathbf{M})^{-1}, \quad (4)$$

$$\mathbf{R} = \text{diag}(\mathbf{M} \cdot \vec{\mathbf{1}})^{-1} \cdot \mathbf{M}, \quad (5)$$

where the function  $\text{diag}(\vec{x})$  takes the elements of the vector  $\vec{x}$  and puts them into the diagonal of a square matrix.

## 2.1 The Product Space

We note that we can transition from products to countries, and back from countries to products. thus, the transition matrix that takes me from products to products is the multiplication of  $\mathbf{R}^T$  and  $\mathbf{Q}$ :

$$\vec{\pi}(t+2) = \mathbf{R}^T \cdot \vec{\gamma}(t+1), \quad (\text{transposing eq. (3)}) \quad (6)$$

$$= \mathbf{R}^T \cdot \mathbf{Q} \cdot \vec{\pi}(t), \quad (\text{using eq. (2)}). \quad (7)$$

Hence, the transition left-stochastic matrix that takes me from products to products is

$$\mathbf{S}_{\mathbf{pp}} = \mathbf{R}^T \cdot \mathbf{Q}.$$

This matrix *is* the Product Space. Note, for instance, that the *dominant right-eigenvector* (corresponding to the eigenvalue  $\lambda_1 = 1$ ) of  $\mathbf{S}_{\mathbf{pp}}$  is the stationary distribution of the random walker at  $t \rightarrow \infty$ , such that  $\vec{\pi}_\infty = \mathbf{S}_{\mathbf{pp}} \vec{\pi}_\infty$  (sometimes called the perron-vector).

Note that the Country Space can be constructed in a similar way, yielding

$$\mathbf{S}_{\mathbf{cc}} = \mathbf{Q} \cdot \mathbf{R}^T.$$

### 2.1.1 Commentary on Complexity Indices

Notice that our complexity indices are defined such that:

$$\vec{pci}^T \cdot \mathbf{S}_{\mathbf{pp}} = \lambda_2^{(products)} \vec{pci}^T, \quad (8)$$

and

$$\vec{eci}^T \cdot \mathbf{S}_{cc} = \lambda_2^{(countries)} \vec{eci}^T, \quad (9)$$

where  $\lambda_2^{(products)}$  and  $\lambda_2^{(countries)}$  are the *sub-dominant* eigenvalues of the product and country spaces, respectively.

### 3 The solution to our problem

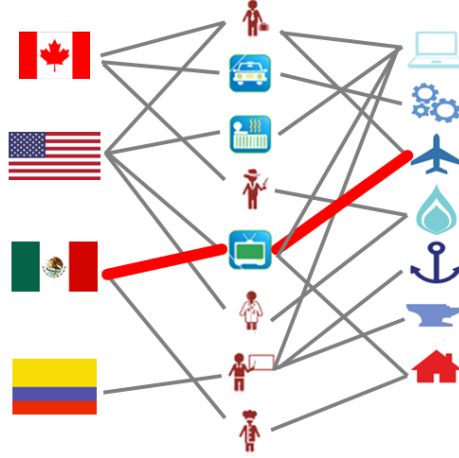


Figure 1: The networks that connect countries with capabilities, and capabilities to products, can be re-expressed as stochastic matrices. Since the product of stochastic matrices is still a stochastic matrix, factorization may be possible.

#### 3.1 Expressing our matrices properly

Assume now that between countries and products there is an intermediate step made of capabilities (see fig. 1). A random walker in this expanded

network will transition, for example, from products to capabilities, and from capabilities to countries. The product space, in this case, will be defined by a transition of the form:  $p' \rightarrow a \rightarrow c \rightarrow a \rightarrow p$ , where  $a$  denotes a single capability  $a$  (with  $a = 1, \dots, N_\alpha$ ). The product space is thus a multiplication of four transition matrices.

The transition matrix  $\mathbf{R}$  from countries to products can now be expressed in terms of  $\mathbf{C}_{\mathbf{ca}}$  and  $\mathbf{P}_{\mathbf{ap}}$  matrices as:

$$\mathbf{R} = \left( \text{diag}(\mathbf{C}_{\mathbf{ca}} \cdot \vec{\mathbf{1}})^{-1} \cdot \mathbf{C}_{\mathbf{ca}} \right) \cdot \left( \text{diag}(\mathbf{P}_{\mathbf{ap}} \cdot \vec{\mathbf{1}})^{-1} \cdot \mathbf{P}_{\mathbf{ap}} \right) \quad (10)$$

$$= (\mathbf{C} \Delta_{\mathbf{c}}^{-1} \cdot \mathbf{C}_{\mathbf{ca}}) \cdot (\mathbf{P} \Delta_{\mathbf{a}}^{-1} \cdot \mathbf{P}_{\mathbf{ap}}), \quad (11)$$

where  $\mathbf{C} \Delta_{\mathbf{c}}$  is a square diagonal matrix of dimensions  $N_c \times N_c$  whose diagonal elements are the number of capabilities that each country has, and similarly,  $\mathbf{P} \Delta_{\mathbf{a}}$  is a square diagonal matrix of dimensions  $N_\alpha \times N_\alpha$  whose diagonal elements are the number of products that make us of each of the capabilities.

Conversely, the transition matrix from products to countries is

$$\mathbf{Q} = \left( \mathbf{C}_{\mathbf{ca}} \cdot \text{diag}(\vec{\mathbf{1}}^T \cdot \mathbf{C}_{\mathbf{ca}})^{-1} \right) \cdot \left( \mathbf{P}_{\mathbf{ap}} \cdot \text{diag}(\vec{\mathbf{1}}^T \cdot \mathbf{P}_{\mathbf{ap}})^{-1} \right) \quad (12)$$

$$= (\mathbf{C}_{\mathbf{ca}} \cdot \mathbf{C} \Delta_{\mathbf{a}}^{-1}) \cdot (\mathbf{P}_{\mathbf{ap}} \cdot \mathbf{P} \Delta_{\mathbf{p}}^{-1}), \quad (13)$$

where  $\mathbf{C} \Delta_{\mathbf{a}}$  is a square diagonal matrix of dimensions  $N_a \times N_a$  whose diagonal elements are the number of countries using each capability, and similarly,  $\mathbf{P} \Delta_{\mathbf{p}}$  is a square diagonal matrix of dimensions  $N_p \times N_p$  whose diagonal elements are the number of capabilities per product.

### 3.2 How $\mathbf{M}_{\mathbf{cp}}$ is factorized

Using all the above equations:

$$\mathbf{M} = (\mathbf{C}_{\mathbf{ca}} \cdot \mathbf{C} \Delta_{\mathbf{a}}^{-1}) \cdot (\mathbf{P}_{\mathbf{ap}} \cdot \mathbf{P} \Delta_{\mathbf{p}}^{-1}) \cdot \text{diag}(\vec{\mathbf{1}}^T \cdot \mathbf{M}), \quad (14)$$

or,

$$\mathbf{M} = \text{diag}(\mathbf{M} \cdot \vec{\mathbf{1}}) \cdot (\mathbf{C} \Delta_{\mathbf{c}}^{-1} \cdot \mathbf{C}_{\mathbf{ca}}) \cdot (\mathbf{P} \Delta_{\mathbf{a}}^{-1} \cdot \mathbf{P}_{\mathbf{ap}}). \quad (15)$$

### 3.3 The number of capabilities per country and per product

Hausmann & Hidalgo (2011) provide some functional forms about the average relationship between the number of capabilities a country has, and its *diversity* (i.e., the number of products it has). The same framework relates the number of capabilities per product with its *ubiquity* (i.e., the number of countries that produce it). On average, we expect

$$k_{c,0}^a = N_a + \frac{1}{q} \ln(d_c), \quad (16)$$

where  $d_c$  is the fraction of all products that country  $c$  produces (i.e., diversity normalized between 0 and 1),  $k_{c,0}^a$  is the number of capabilities that country  $c$  has, and  $q$  is the average fraction of capabilities per product. And

$$k_{p,0}^a = \frac{\ln(u_p)}{\ln(r)}, \quad (17)$$

where  $u_p$  is the fraction of all countries that produce  $p$  (i.e., ubiquity normalized between 0 and 1),  $k_{p,0}^a$  is the number of capabilities required by each product, and  $r$  is the average number of capabilities per country.

### 3.4 One way

One way to identify  $\mathbf{C}_{\mathbf{ca}}$  and  $\mathbf{P}_{\mathbf{ap}}$  is by trying to find some way of separating in three the following two equations, and taking the left- and right-most matrices.

$$\mathbf{M} \cdot f(k_{p,0}^a, u_p) = \mathbf{C}_{\mathbf{ca}} \cdot \mathbf{C} \mathbf{\Delta}_{\mathbf{a}}^{-1} \cdot \mathbf{P}_{\mathbf{ap}}, \quad (18)$$

or,

$$g(k_{c,0}^a, d_c) \cdot \mathbf{M} = \mathbf{C}_{\mathbf{ca}} \cdot \mathbf{P} \mathbf{\Delta}_{\mathbf{a}}^{-1} \cdot \mathbf{P}_{\mathbf{ap}}. \quad (19)$$

### 3.5 The other way

The other way is to hope the following works.

Factorize (non-negatively)  $\mathbf{R}$  and  $\mathbf{Q}$  such that:

$$\begin{aligned}\mathbf{R} &= (\mathbf{C}\Delta_{\mathbf{c}}^{-1} \cdot \mathbf{C}_{\mathbf{ca}}) \cdot (\mathbf{P}\Delta_{\mathbf{a}}^{-1} \cdot \mathbf{P}_{\mathbf{ap}}), \\ &\approx \hat{\mathbf{A}} \cdot \hat{\mathbf{B}}.\end{aligned}\tag{20}$$

Similarly

$$\begin{aligned}\mathbf{Q} &= (\mathbf{C}_{\mathbf{ca}} \cdot \mathbf{C}\Delta_{\mathbf{a}}^{-1}) \cdot (\mathbf{P}_{\mathbf{ap}} \cdot \mathbf{P}\Delta_{\mathbf{p}}^{-1}), \\ &\approx \hat{\mathbf{C}} \cdot \hat{\mathbf{D}}.\end{aligned}\tag{21}$$

Since we can guess  $\mathbf{C}\Delta_{\mathbf{c}}$  from eq. (16), and  $\mathbf{P}\Delta_{\mathbf{p}}$  from eq. (17), then we estimate our desired matrices as  $\mathbf{C}_{\mathbf{ca}} \approx \mathbf{C}\Delta_{\mathbf{c}} \cdot \hat{\mathbf{A}}$  and  $\mathbf{P}_{\mathbf{ap}} \approx \hat{\mathbf{D}} \cdot \mathbf{P}\Delta_{\mathbf{p}}$ .

## 4 Some problems for the ‘Delta’ matrices

(I’m not sure of all this... but it’s what I think we discussed last times)

We said we could fill in the diagonal elements of  $\mathbf{C}\Delta_{\mathbf{c}}$  and  $\mathbf{P}\Delta_{\mathbf{p}}$  which are, respectively, the number of capabilities per country, and per product. In other words, their complexities.

## 5 Product Space

- A similarity matrix between  $p$  and  $p'$ .
- Hidalgo et al. (2007):

$$\phi(p, p') = \min\{\Pr(p|p'), \Pr(p'|p)\} \quad (22)$$

- Stock et al.:

$$\phi(p, p') = \text{cor}(r\vec{c}a(p), r\vec{c}a(p')) \quad (23)$$

But let's take for a moment the original product space and “un-pack” it mathematically.

$$\begin{aligned} \Pr(p|p') &= \frac{\Pr(p, p')}{\Pr(p')} \\ &= \frac{N(p, p')/N_c}{N(p')/N_c} \\ &= \frac{\sum_c M_{c,p} M_{c,p'}}{\sum_c M_{c,p'}} \\ &= \frac{\sum_c M_{c,p} M_{c,p'}}{u(p')}. \end{aligned} \quad (24)$$

Symmetric matrix:

$$\Phi = \min \left\{ \mathbf{U}^{-1} \cdot \mathbf{M}^T \cdot \mathbf{M}, \quad \mathbf{M}^T \cdot \mathbf{M} \cdot \mathbf{U}^{-1} \right\}.$$

Other options

- $\Phi = \mathbf{M}^T \mathbf{M} / N_c$  (symmetric [joint freq.])
- $\Phi = \mathbf{M}^T \cdot \mathbf{D}^{-1} \cdot \mathbf{M}$  (symmetric [“diversity normalized” joint freq.])
- $\Phi = (\mathbf{D}^{-1} \mathbf{M})^T \cdot (\mathbf{D}^{-1} \mathbf{M})$  (symmetric [“averaged probabilities”])
- $\Phi = \mathbf{M}^T \cdot \mathbf{D}^{-2} \cdot \mathbf{M}$  (symmetric [“averaged probabilities”])
- $\Phi = \mathbf{U}^{-1} \cdot \mathbf{M}^T \cdot \mathbf{M} \cdot \mathbf{U}^{-1}$  (symmetric [mutual information])



## 6 Density regressions

Constructing a density:

$$\begin{aligned} M_{c,p}(t+1) &= \sum_{p'} M_{c,p'}(t) \frac{\phi(p', p)}{\sum_{p''} \phi(p'', p)} \\ &= \sum_{p'} M_{c,p'}(t) P(p', p) \quad \text{where } \sum_{p'} P(p', p) = 1. \end{aligned} \quad (25)$$

It is a weighted average. In matrix form

$$\mathbf{M}(t+1) = \mathbf{M}(t) \cdot \mathbf{P}, \quad (26)$$

where the elements of a column in  $\mathbf{P}$  add up to 1. This is what is called a “stochastic matrix”.

A couple of comments:

**Multiplying a row vector on the left of  $\mathbf{P}$ :** Computes weighted averages,

$$\vec{x}(t+1)^T = \vec{x}(t)^T \cdot \mathbf{P}.$$

“Consensus dynamics”

**Multiplying a column vector on the right of  $\mathbf{P}$ :** Propagates/diffuses the values,

$$\vec{n}(t+1) = \mathbf{P} \cdot \vec{n}(t).$$

“Diffusion dynamics”

Other comments:

**Product-based density**

$$\mathbf{M}(t+1) = \mathbf{M}(t) \cdot \mathbf{P}.$$

**Country-based density**

$$\mathbf{M}(t+1) = \mathbf{C} \cdot \mathbf{M}(t).$$

**Together**

$$\mathbf{M}(t+1) = \mathbf{M}(t) \cdot \mathbf{P} + \mathbf{C} \cdot \mathbf{M}(t).$$

Maybe there is an external driving force:

$$\mathbf{M}(t+1) = \mathbf{M}(t) \cdot \mathbf{P} + \mathbf{C} \cdot \mathbf{M}(t) + \mathbf{U}(t), \quad (27)$$

Low dimensional representation.

Given the following left-eigenvectors

$$\vec{\psi}_k^T \cdot \mathbf{P} = \lambda_k \vec{\psi}_k^T, \quad (28)$$

where  $\lambda_1 = 1 \geq |\lambda_k| \geq 0$ .

$$\begin{aligned} \vec{x}(t+1)^T &= \vec{x}(t)^T \cdot \mathbf{P} \\ &= \left( \sum_k a_k \vec{\psi}_k^T \right) \cdot \mathbf{P} \\ &= \sum_k a_k \vec{\psi}_k^T \cdot \mathbf{P} \\ &= \sum_k \lambda_k a_k \vec{\psi}_k^T \\ &\approx \lambda_1 a_1 \vec{\psi}_1^T + \lambda_2 a_2 \vec{\psi}_2^T. \end{aligned}$$

## 7 ECI and the method of reflections