PRO
MATHEMATICA

VOLUMEN XXVIII / No. 55 / 2014

Oswaldo Velásquez Castañon
IMCA, *Instituto de Matemática y Ciencias Afines*
ovelasquez@imca.edu.pe

Abel Cadenillas
University of Alberta, Canadá
acadenil@math.ualberta.ca

Fernando Torres Orihuela
UNICAMP, *Universidade Estadual de Campinas*, Brasil
ftorres@ime.unicamp.br

PRO MATHEMATICA

La revista PRO MATHEMATICA, fundada en el año 1987, es una publicación del Departamento de Ciencias, Sección Matemáticas, de la Pontificia Universidad Católica del Perú, PUCP. Esta revista acoge artículos de investigación de alto estándar tanto en exposición como en contenido sea en matemáticas puras, matemáticas aplicadas o estadística. Anualmente se edita un volumen compuesto de dos números. Todos los artículos pasan por riguroso proceso de arbitraje. PRO MATHEMATICA se encuentra actualmente indexada en el catálogo *Latindex*. Desde el año 2014 la distribución será exclusivamente electrónica.

Founded in 1987, PRO MATHEMATICA is a regular journal published by the Department of Sciences, Mathematics Section, of Pontificia Universidad Católica del Perú, PUCP. This journal shelters research papers of high standards, both in content as in exposition, in Pure and Applied Mathematics and Statistics. A volume composed by two numbers is published yearly. All articles are peer refereed. PRO MATHEMATICA nowadays is indexed in the *Latindex* catalog. From 2014 on, the distribution would be exclusively by electronic means.

Composición de textos
Huarcaya Huarcaya Karol José María

PRO MATHEMATICA

VOLUMEN XXVIII / Nº 55 / 2014

CONTENIDO

Involuciones, trivoluciones y foliaciones Galois

***A. Beltrán*[1], *M. Falla*[2] *y D. Marín*[3]**

Febrero, 2014

Resumen

En el presente trabajo introducimos la noción de foliaciones Galois sobre $\mathbb{P}^2_{\mathbb{C}}$, definidas como aquellas cuya aplicación de Gauss restringida a un abierto Zariski es un recubrimiento Galois. Asimismo, presentamos algunos ejemplos y un criterio para identificar este tipo de foliaciones.

MSC(2010): 53A60.

Palabras Clave: Foliaciones, webs.

[1] *Sección Matemáticas, Departamento de Ciencias, PUCP*
PUCP-DGI-2013-0014.
[2] *Departamento de Análise-IM, UFF, Brasil.*
[3] *Departament de Matemàtiques, UAB, España.*

1. Involuciones

Una transformación racional en el plano proyectivo complejo es una aplicación de la forma

$$\begin{array}{rccc} f: & \mathbb{P}^2_{\mathbb{C}} & \dashrightarrow & \mathbb{P}^2_{\mathbb{C}} \\ & [x,y,z] & \longmapsto & [A_1(x,y,z), A_2(x,y,z), A_3(x,y,z)], \end{array}$$

donde A_i, $i = 1, 2, 3$, son polinomios homogéneos del mismo grado sin factor común. Una transformación birracional $f : \mathbb{P}^2_{\mathbb{C}} \dashrightarrow \mathbb{P}^2_{\mathbb{C}}$ es una transformación racional que admite una inversa racional g: $f \circ g = g \circ f = \mathrm{Id}$. Denotamos por $\mathrm{Bir}(\mathbb{P}^2_{\mathbb{C}})$ al grupo de transformaciones birracionales del plano, llamado **grupo de Cremona**. Sea $f = [A_1, A_2, A_3]$ una transformación birracional. Definimos el grado de f como el grado común de los A_i, esto es, $\deg f = \deg A_i$. Asimismo, el **conjunto de indeterminación** $\mathrm{Ind}(f)$ y el **conjunto excepcional** $\mathrm{Exc}(f)$ son definidos respectivamente por

$$\begin{aligned} \mathrm{Ind}(f) &= \{p \in \mathbb{P}^2_{\mathbb{C}} : A_1(p) = A_2(p) = A_3(p) = 0\}, \\ \mathrm{Exc}(f) &= \{p \in \mathbb{P}^2_{\mathbb{C}} : \det Jac(f(p)) = 0\}. \end{aligned}$$

Ejemplo 1.1. El automorfismo $f : \mathbb{P}^2_{\mathbb{C}} \dashrightarrow \mathbb{P}^2_{\mathbb{C}}$ dado por

$$f[x,y,z] = [a_{11}x + a_{12}y + a_{13}z, a_{21}x + a_{22}y + a_{23}z, a_{31}x + a_{32}y + a_{33}z],$$

con $\det(a_{ij}) \neq 0$, es una transformación birracional de grado 1. Para ésta se tiene $\mathrm{Ind}(f) = \mathrm{Exc}(f) = \emptyset$. Estas tranformaciones birracionales son llamadas de **tipo proyectivo**.

Ejemplo 1.2. La transformación $f[x,y,z] = [xy, z^2, yz]$ es birracional y cumple

$$\begin{aligned} \mathrm{Ind}(f) &= \{[1,0,0], [0,1,0]\}, \\ \mathrm{Exc}(f) &= \{y = 0\} \cup \{z = 0\}, \\ f^2 &= \mathrm{Id}. \end{aligned}$$

Ejemplo 1.3. La transformación $f[x,y,z] = [yz, xz, xy]$ es birracional y cumple

$$\begin{aligned} \mathrm{Ind}(f) &= \{[1,0,0],[0,1,0],[0,0,1]\},\\ \mathrm{Exc}(f) &= \{x=0\}\cup\{y=0\}\cup\{z=0\},\\ f^2 &= \mathrm{Id}. \end{aligned}$$

Esta transformación es llamada **de Cremona** y es conjugada por una aplicación birracional a una de Jonquières (ver Teorema 1.5). Notemos que f preserva las fibraciones racionales $\dfrac{y}{x} = c_{yx}$, $\dfrac{x}{z} = c_{xz}$, $\dfrac{y}{z} = c_{yz}$, con c_{yx}, c_{xz}, c_{yz} constantes.

Definición 1.4. Una **involución birracional** es una aplicación birracional $f \in \mathrm{Bir}(\mathbb{P}^2_{\mathbb{C}})$ que satisface $f^2 = \mathrm{Id}$.

1.1. Involuciones de Geiser

Sean $p_1, \dots p_7 \in \mathbb{P}^2_{\mathbb{C}}$ siete puntos en $\mathbb{P}^2_{\mathbb{C}}$ en posición general, y denotemos por L el sistema lineal de curvas cúbicas que pasan por los puntos p_i, en símbolos ponemos

$$L = \{C \text{ cúbica} : \ p_i \in C\} \cong \mathbb{P}^2_{\mathbb{C}},$$

donde una cúbica está dada por polinomios homogéneos de grado 3 en tres variables. Sea $p \in \mathbb{P}^2_{\mathbb{C}}$ un punto genérico. Entonces el conjunto

$$L_p = \{C \in L : \ p \in C\} \subset \mathbb{P}^2_{\mathbb{C}}$$

es un **lápiz de cúbicas**. Por el teorema de Bezout L_p tiene 9 puntos base: $p, p_1, \dots p_7, I_G(p)$. De esta manera queda definida

$$\begin{array}{rccc} I_G: & \mathbb{P}^2_{\mathbb{C}} & \dashrightarrow & \mathbb{P}^2_{\mathbb{C}} \\ & p & \longmapsto & I_G(p), \end{array}$$

donde $I_G(p)$ es el noveno punto base del lápiz L_p. Una técnica estandar muestra que I_G es una involución birracional, llamada **involución de Geiser** (cf. [4]).

Existen otras construcciones geométricas que dan lugar a nuevas involuciones birracionales, por ejemplo las llamadas de Bertini y de Jonquières, que no discutiremos en este trabajo. Para mayores detalles referimos al lector a [4, 1].

Teorema 1.5 ([1])**.** *Una involución birracional de $\mathbb{P}^2_{\mathbb{C}}$ es conjugada a una de las siguientes involuciones: proyectiva, de Jonquières, de Bertini, de Geiser.* □

2. Foliaciones de grado 2

Sea $\mathcal{F}$ una foliación holomorfa sobre $\mathbb{P}^2_{\mathbb{C}}$ de grado d, con conjunto singular $\Sigma_{\mathcal{F}}$, y sea ℓ una recta genérica. El **grado de** $\mathcal{F}$ es por definición el número de puntos de tangencia entre $\mathcal{F}$ y ℓ, contando multiplicidades. Consideremos una foliación $\mathcal{F}$ de grado 2 en $\mathbb{P}^2_{\mathbb{C}}$, y $p \in \mathbb{P}^2_{\mathbb{C}}$, un punto genérico. Por definicion $\mathcal{F}$ y $T_p\mathcal{F}$ tienen dos puntos de contacto, uno de ellos es p y al otro lo denotaremos por $I_{\mathcal{F}}(p)$. Para precisar esta definición supongamos que $\mathcal{F}$ es definida por un campo de la forma $X(x,y) = A(x,y)\frac{\partial}{\partial x} + B(x,y)\frac{\partial}{\partial y}$ en la carta afín (x,y) de $\mathbb{P}^2_{\mathbb{C}}$. Entonces $q = I_{\mathcal{F}}(p)$ es el punto $p + t(A,B)$, donde t es el único parámetro no nulo donde $X(p)$ es colineal con $X(p + t(A,B))$. Es claro que $I_{\mathcal{F}}$ es una involución birracional, llamada **involución asociada a** $\mathcal{F}$ (cf. [4]).

Teorema 2.1 (Cerveau-Deserti [4])**.** *Sean $p_1, \ldots, p_7$ puntos de $\mathbb{P}^2_{\mathbb{C}}$ en posición general. Sea $\mathcal{F}$ una foliación de grado 2 sobre $\mathbb{P}^2_{\mathbb{C}}$ cuyo conjunto singular $\Sigma_{\mathcal{F}}$ está formado por los 7 puntos dados. Entonces $I_{\mathcal{F}} = I_G$. En particular, la involución asociada a una foliación genérica de grado 2 es una involución de Geiser.* □

Tradicionalmente las involuciones de Geiser son definidas a partir de lápices de cúbicas. Sin embargo, es un problema de interés la construcción explícita de tales involuciones. Es de ahí de donde salta la importancia del teorema anterior.

Ejemplo 2.2. Sea $\mathcal{F}$ la foliación dada por el campo $X = (x^3 - y^2)\frac{\partial}{\partial x} + (x^2y - 1)\frac{\partial}{\partial y}$, en coordenadas afines. Teniendo en mente la construcción anterior deducimos que la involución asociada a $\mathcal{F}$ es $I_{\mathcal{F}}[x, y, z] = [A_1, A_2, A_3]$, donde

$$\begin{aligned} A_1 &= xy^7 + 3x^5y^2z - x^8 - 5x^2y^4z^2 + 2y^3z^5 + x^3yz^4 - xz^7, \\ A_2 &= 3xy^5z^2 + 2x^5z^3 - x^7y - 5x^2y^2z^4 + x^4y^3z + yz^7 - y^8, \\ A_3 &= xy^4z^3 - 5x^4y^2z^2 - y^7z + 2x^3y^5 + 3x^2yz^5 - z^8 + x^7z. \end{aligned}$$

Además, el conjunto de indeterminación está dado por

$$\mathrm{Ind}(I_{\mathcal{F}}) = \{[\xi^j, \xi^{-2j}, 1] : \ j = 0, \ldots, 6, \text{ donde } \xi^7 = 1\} = \Sigma_{\mathcal{F}}.$$

Estos puntos se encuentran en posición general, y por lo tanto estamos ante una involución de Geiser.

3. Foliaciones Galois de grado $d \geq 3$

Sea $\mathcal{F}$ una foliación de grado 3. Esto significa que una recta genérica es tangente a $\mathcal{F}$ en tres puntos. Ahora el juego trata de averiguar si es posible construir una transformación birracional que permute dichos puntos. La respuesta suele ser negativa puesto que una transformación que intercambie estos puntos será genéricamente multivaluada. Sin embargo, Cerveau y Deserti en [4] proporcionan un criterio para asegurar cuándo dicha transformación es birracional, y en [3] este criterio se generaliza para foliaciones sobre $\mathbb{P}^2_{\mathbb{C}}$ de grado d.

Definición 3.1. Una aplicación $\tau \in \mathrm{Bir}(\mathbb{P}^2_{\mathbb{C}})$ es llamada **trivolución** si satisface $\tau^3 = Id$.

Es claro que de existir una aplicación birracional τ que intercambia cíclicamente las tangencias de una foliación de grado 3, esta aplicación será una trivolución.

La **aplicación de Gauss asociada** a $\mathcal{F}$ es la aplicación racional

$$\begin{array}{rccc} \mathcal{G}_{\mathcal{F}} : & \mathbb{P}^2_{\mathbb{C}} & \dashrightarrow & \check{\mathbb{P}}^2_{\mathbb{C}} \\ & p & \longmapsto & T_p\mathcal{F}, \end{array}$$

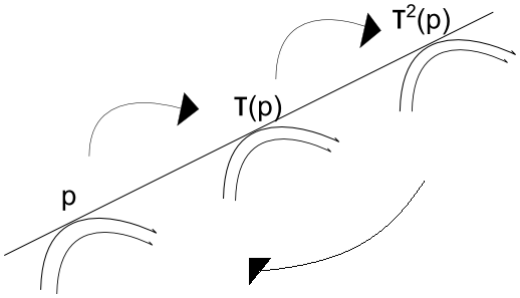

Figura 1: Trivolución asociada a una foliación

donde $T_p\mathcal{F}$ denota la recta tangente a $\mathcal{F}$ en un punto regular p de $\mathcal{F}$.

Si la foliación $\mathcal{F}$ es dada por la forma $\omega = P(x,y,z)dx+Q(x,y,z)dy+R(x,y,z)dz$, entonces la aplicación de Gauss queda materializada por

$$\mathcal{G}_{\mathcal{F}}(p) = [P(p), Q(p), R(p)].$$

Notemos que la propiedad de τ de intercambiar las tangencias equivale a que se cumpla $\mathcal{G}_{\mathcal{F}} \circ \tau = \mathcal{G}_{\mathcal{F}}$.

Sea $\rho : X \longrightarrow B$ un recubrimiento de grado d entre espacios conexos. Fijamos un punto $b \in B$ y su fibra $F = \rho^{-1}(b) = \{p_1, \ldots, p_d\}$. También, consideremos el grupo $D = \{\tau : X \longrightarrow X : \rho \circ \tau = \rho\}$ de automorfismos de recubrimiento, que actúa a la izquierda sobre F. La representación de monodromía $\mu : \pi_1(B,x) \longrightarrow \mathrm{Aut}(F)$ actúa a la derecha sobre F, y a su imagen la denotaremos por M. Se cumple que D actúa libremente sobre F, mientras M lo hace transitivamente sobre F. Un simple conteo conduce a $|D| \leq d$ y $|M| \geq d$. Al identificar $F \simeq \{1, \ldots d\}$, podemos considerar D y M como subgrupos del grupo simétrico S_d. El siguiente resultado conocido brinda una relación especial entre los grupos D y M; ver por ejemplo [6] y [8].

Teorema 3.2. *Si $\rho : X \longrightarrow B$ es un recubrimiento de grado d, las siguientes afirmaciones*

- $\rho_*(\pi_1(X,x)) = \ker \mu$*;*
- $\rho_*(\pi_1(X,x))$ *es un subgrupo normal de* $\pi_1(B,b)$*;*
- D *actúa transitivamente sobre la fibra* F*;*
- $|D| = d$*;*
- M *actúa libremente sobre* F*;*
- $|M| = d$*;*
- $D \cong M$

son equivalentes. □

Definición 3.3. Decimos que ρ es **Galois** si se cumple cualquiera de las afirmaciones enumeradas en el teorema 3.2.

Definición 3.4. Un $k-$**web** $\mathcal{W}$ **de codimensión uno** sobre una variedad compleja S está dado por una cobertura abierta $\{U_i\}$ junto con $k-$formas simétricas $\omega_i \in \mathrm{Sym}^k_S(U_i)$ que verifican las siguientes condiciones:

i) para cada intersección no vacía $U_i \cap U_j$ existe una función no nula $g_{ij} \in \mathcal{O}^*(U_i \cap U_j)$ tal que $\omega_i = g_{ij}\omega_j$;

ii) el conjunto de ceros $\mathrm{Sing}(\omega_i)$ de ω_i tiene codimensión al menos dos;

iii) el germen de ω_i en cada punto genérico de U_i es un producto de $k-$gérmenes de $1-$formas integrables que no son colineales dos a dos.

El subconjunto de S donde la condición (*iii*) falla es llamado **discriminante del web** y se denota por $\Delta(\mathcal{W})$. **El conjunto singular** $\Sigma_{\mathcal{W}}$ de $\mathcal{W}$ es definido por $\Sigma_{\mathcal{W}} \cap U_i = \mathrm{Sing}(\omega_i)$ y se encuentra contenido en $\Delta(\mathcal{W})$. Existe una representación de monodromía $\mu : \pi_1(S \backslash \Delta(\mathcal{W})) \longrightarrow \mathfrak{S}_k$ de $\mathcal{W}$ que determina los subwebs irreducibles de $\mathcal{W}$ y cuya trivialidad es equivalente a la descomponibilidad del web $\mathcal{W}$ (ver también definición 3.11).

Sea $\mathcal{F}$ una foliación sobre $\mathbb{P}^2_{\mathbb{C}}$ y $\mathcal{G}_{\mathcal{F}} : \mathbb{P}^2_{\mathbb{C}} \dashrightarrow \check{\mathbb{P}}^2_{\mathbb{C}}$ la aplicación de Gauss asociada, representada por el morfismo $\mathcal{G}_{\mathcal{F},U} : U \longrightarrow \check{\mathbb{P}}^2_{\mathbb{C}}$ definido sobre un subconjunto abierto maximal denso Zariski de $\mathbb{P}^2_{\mathbb{C}}$. Si $\mathcal{W}$ es un

$k-$web sobre $\check{\mathbb{P}}^2_{\mathbb{C}}$ dado por el par de colecciones $\{V_i, \eta_i\}$ que verifican la definición 3.4, entonces podemos definir un web $\mathcal{G}^*_{\mathcal{F}}(\mathcal{W})$ sobre $\mathbb{P}^2_{\mathbb{C}}$ por la colección $\{U_i = \mathcal{G}_{\mathcal{F}}^{-1}(V_i), \omega_i\}$, donde $\omega_i = \mathcal{G}^*_{\mathcal{F},U}\eta_i$. El web $\mathcal{G}^*_{\mathcal{F}}\mathcal{W}$ es llamado **pull-back** o **imagen inversa** de $\mathcal{W}$ por $\mathcal{G}_{\mathcal{F}}$.

El **grado de un** $k-$**web** $\mathcal{W}$ sobre $\mathbb{P}^2_{\mathbb{C}}$ está dado por la suma de la cantidad de puntos de tangencia que tiene una recta genérica no invariante con las foliaciones que componen el web.

Denotamos por $\mathbb{W}(k,d)$ al conjunto de $k-$webs de grado d sobre $\mathbb{P}^2_{\mathbb{C}}$ Definimos la aplicación

$$\text{Leg} : \mathbb{W}(k,d) \longrightarrow \mathbb{W}(d,k),$$

llamada **transformado de Legendre o web dual** de $\mathcal{W}$, de la siguiente manera. Sea $\mathcal{W} \in \mathbb{W}(k,d)$ fijo y $\ell \in \check{\mathbb{P}}^2_{\mathbb{C}}$, visto como una recta en $\mathbb{P}^2_{\mathbb{C}}$. Si ℓ es genérica entonces se tienen d puntos de tangencia distintos $p_1, \ldots, p_d$ con las hojas de $\mathcal{W}$. Por supuesto, podemos pensar estos puntos como rectas sobre $\check{\mathbb{P}}^2_{\mathbb{C}}$ que pasan por ℓ. Variando el punto ℓ obtenemos d direcciones diferentes sobre $\check{\mathbb{P}}^2_{\mathbb{C}}$ que definen un $d-$web $\text{Leg}\,\mathcal{W}$. De manera rutinaria se verifica que el grado de $\text{Leg}\,\mathcal{W}$ es k. Las hojas de $\text{Leg}\,\mathcal{W}$ son las curvas duales de las hojas de $\mathcal{W}$ puesto que ellas son tangentes a las direcciones que definen $\text{Leg}\,\mathcal{W}$. En particular, si $\mathcal{F}$ es una foliación de grado d sobre $\mathbb{P}^2_{\mathbb{C}}$, entonces $\text{Leg}\,\mathcal{F}$ es un $d-$web de grado 1.

Con la notacion $\check{\Lambda}_{\mathcal{F}} = \mathcal{G}_{\mathcal{F}}(\mathcal{I}_{\mathcal{F}}) \subset \check{\mathbb{P}}^2_{\mathbb{C}}$, $\Lambda_{\mathcal{F}} = \mathcal{G}_{\mathcal{F}}^{-1}(\mathcal{G}_{\mathcal{F}}(\mathcal{I}_{\mathcal{F}}))$ y $\mathcal{G}^{\Lambda}_{\mathcal{F}} = \mathcal{G}_{\mathcal{F}}\Big|_{\mathbb{P}^2_{\mathbb{C}}\setminus\Lambda_{\mathcal{F}}}$ tenemos el siguiente resultado.

Teorema 3.5. *Sea $\mathcal{F}$ una foliación de grado d sobre $\mathbb{P}^2_{\mathbb{C}}$. Entonces la restricción de la aplicación de Gauss a $\mathbb{P}^2_{\mathbb{C}}\setminus\Lambda_{\mathcal{F}}$, pongamos $\mathcal{G}^{\Lambda}_{\mathcal{F}} : \mathbb{P}^2_{\mathbb{C}}\setminus\Lambda_{\mathcal{F}} \longrightarrow \check{\mathbb{P}}^2_{\mathbb{C}}\setminus\check{\Lambda}_{\mathcal{F}}$, es un recubrimiento de grado d. Más aún, el grupo de aplicaciones de recubrimiento $D_{\mathcal{G}_{\mathcal{F}}}$ es obtenido por restricción a $\mathbb{P}^2_{\mathbb{C}}\setminus\Lambda_{\mathcal{F}}$ del grupo $D_{\mathcal{F}} = \{\tau \in Bir(\mathbb{P}^2_{\mathbb{C}}) : \mathcal{G}_{\mathcal{F}} \circ \tau = \mathcal{G}_{\mathcal{F}}\}$. Además, si $\mathcal{W}_{\mathcal{F}} = Leg\,\mathcal{F}$ es el transformado de Legendre de $\mathcal{F}$, con discriminante $\check{\Delta}_{\mathcal{F}}$, entonces la monodromía de $\mathcal{G}^{\Lambda}_{\mathcal{F}}$ está dada por la composición*

$$\pi_1(\check{\mathbb{P}}^2_{\mathbb{C}}\setminus\check{\Lambda}_{\mathcal{F}}) \twoheadrightarrow \pi_1(\check{\mathbb{P}}^2_{\mathbb{C}}\setminus\check{\Delta}_{\mathcal{F}}) \longrightarrow \mathfrak{S}_d\ ,$$

donde el primera aplicación es inducida por la inclusión natural.

Prueba. Ver [3, Teorema 3.11]. □

Definición 3.6. Decimos que una foliación $\mathcal{F}$ sobre $\mathbb{P}^2_{\mathbb{C}}$ es **Galois** si la restricción $\mathcal{G}^{\Lambda}_{\mathcal{F}}$ de la aplicación de Gauss es Galois.

El criterio descrito en [4] para determinar cuándo una foliación es Galois puede reformularse de la siguiente manera.

Proposición 3.7. *Una foliación sobre $\mathbb{P}^2$ definida por el campo vectorial polinomial $A\frac{\partial}{\partial x} + B\frac{\partial}{\partial y}$ es Galois si y solo si el polinomio*

$$P(x,y,t) = det\begin{pmatrix} A(x,y) & A(\,x+tA(x,y),\,y+tB(x,y)) \\ B(x,y) & B(\,x+tA(x,y),\,y+tB(x,y)) \end{pmatrix} \in \mathbb{C}[x,y,t]$$

se descompone totalmente sobre el cuerpo $\mathbb{C}(x,y)$. En tal caso, cada una de sus raíces $t = t(x,y) \in \mathbb{C}[x,y,t]$ determina una transformación de recubrimiento birracional

$$\tau : (x,y) \mapsto (\,x+tA(x,y),\,y+tB(x,y))$$

de $\mathcal{G}_G$. □

Ejemplo 3.8. Toda foliación $\mathcal{F}$ de grado 2 sobre $\mathbb{P}^2_{\mathbb{C}}$ es Galois.

Ejemplo 3.9. Sean $\alpha,\beta,\gamma,\delta \in \mathbb{C}$ tales que $\alpha\delta - \beta\gamma \neq 0$, y $u,v \in \mathbb{C}[x,y]$ polinomios linealmente independientes de grado menor o igual a 1. Entonces la foliación $\mathcal{F}$ de grado d definida por

$$Y = (\alpha u^d + \beta v^d)\frac{\partial}{\partial x} + (\gamma u^d + \delta v^d)\frac{\partial}{\partial y}$$

es Galois y tiene grupo de monodromía cíclico. En efecto, la pendiente del campo vectorial Y tiene la forma $p(x,y) = h(f(x,y))^d$, donde $f(x,y) = \dfrac{u(x,y)}{v(x,y)}$ y $h(z) = \dfrac{\gamma z+\delta}{\alpha z+\beta} \in \mathrm{PSL}_2(\mathbb{C})$. De este modo las raíces del polinomio

$$P(x,y,t) = \det(X(x,y), X(x+t, y+tp(x,y))) \tag{1}$$

son las soluciones de la ecuación $h \circ f^d(x+t, y+tp(x,y)) = h \circ f^d(x,y)$, pues teniendo en cuenta la propiedad $h \in \mathrm{PSL}_2(\mathbb{C})$, de la última relación se obtiene

$$\left(\frac{f(x+t, y+tp(x,y))}{f(x,y)}\right)^d = 1.$$

Si escribimos $\xi = \mathrm{e}^{\frac{2i\pi}{d}}$, las raíces de (1) son las soluciones $t = t(x,y)$ de las d ecuaciones lineales en la variable t dadas por

$$f(x+t, y+tp(x,y)) = \xi^k f(x,y), \quad \text{para } k = 1, \ldots, d.$$

Cada solución $t = t_k(x,y)$ determina una aplicación de recubrimiento

$$\tau_k : (x,y) \longmapsto (x + t_k(x,y), y + t_k(x,y)p(x,y))$$

que verifica $f \circ \tau_k(x,y) = \xi^k f(x,y)$. Luego, con la notación $\tau_k \circ \tau_\ell = \tau_{m(k,\ell)}$, se tiene

$$\xi^{k+\ell} = f \circ \tau_k \circ \tau_\ell(x,y) = f \circ \tau_{m(k,\ell)} = \xi^{m(k,\ell)} f(x,y),$$

es decir, $m(k,\ell) = (k+\ell) \mod d$. De esta manera concluimos la igualdad $D_{\mathcal{F}} = \{\tau_1, \ldots, \tau_d\} = \langle \tau_1 \rangle$. Esto implica las relaciones $M_{\mathcal{F}} \simeq D_{\mathcal{F}} = \mathbb{Z}_d$, pues $|D_{\mathcal{F}}| = d$.

Proposición 3.10. *Sea $\mathcal{F}$ una foliación de grado d sobre $\mathbb{P}^2_{\mathbb{C}}$. Entonces el web $\mathcal{G}^*_{\mathcal{F}}\mathrm{Leg}\,\mathcal{F}$ contiene a la foliación $\tau^*\mathcal{F}$, para cada $\tau \in D_{\mathcal{F}}$.*

Prueba. Sea $\tau \in D_{\mathcal{F}}$ arbitrario y L una hoja de la foliación $\mathcal{F}$. Entonces $\tau^{-1}(L)$ es una hoja de $\tau^*\mathcal{F}$. Del teorema 3.5 se sigue que $\mathcal{G}_{\mathcal{F}}(\tau^{-1}(L)) = \mathcal{G}_{\mathcal{F}}(L)$ es una hoja de $(\mathcal{G}_{\mathcal{F}})_*(\mathcal{F}) = \mathrm{Leg}\,\mathcal{F}$, y por tanto $\tau^{-1}(L)$ resulta una hoja de $\mathcal{G}^*_{\mathcal{F}}\mathrm{Leg}\,\mathcal{F}$. □

Definición 3.11. Un $k-$web W sobre una superficie S es llamado **totalmente descomponible** si sobre S el web W es la superposición de $k-$foliaciones globales. En tal caso, lo denotamos por

$$W = \mathcal{G}_1 \boxtimes \ldots \boxtimes \mathcal{G}_k,$$

donde $\mathcal{G}_i$ es una foliación sobre S.

Ahora pasemos a enunciar y probar nuestro resultado central.

Teorema 3.12. *Una foliación $\mathcal{F}$ de grado d es Galois si y solo si $\mathcal{G}^*_{\mathcal{F}}Leg\,\mathcal{F}$ es completamente descomponible.*

Prueba. De la proposición 3.10 se sigue que $\boxtimes_{\tau\in D_{\mathcal{F}}}\tau^*\mathcal{F}$ es un subweb de $\mathcal{G}^*_{\mathcal{F}}\mathrm{Leg}\mathcal{F}$, en particular se tiene $|D_{\mathcal{F}}|\leq d$. Y como $\mathcal{F}$ es Galois, el teorema 3.5 implica $|D_{\mathcal{F}}|=d$, y por lo tanto el web $\mathcal{G}^*_{\mathcal{F}}\mathrm{Leg}\,\mathcal{F}$ es completamente descomponible.

Recíprocamente, del teorema 3.2 se desprende que $\mathcal{G}^{\Lambda}_{\mathcal{F}}$ es Galois si y solo si se cumple $(\mathcal{G}^{\Lambda}_{\mathcal{F}})_*(\pi_1(\mathbb{P}^2_{\mathbb{C}}\backslash\Lambda_{\mathcal{F}}))\subset\ker\mu_{\mathcal{F}}$, donde $\mu_{\mathcal{F}}$ es la representación de monodromía de $\mathcal{G}^{\Lambda}_{\mathcal{F}}$. Asimismo, del teorema 3.5 se sigue que el diagrama

$$\begin{array}{ccccc}
\pi_1(\mathbb{P}^2_{\mathbb{C}}\backslash\Lambda_{\mathcal{F}}) & \xrightarrow{(\mathcal{G}^{\Lambda}_{\mathcal{F}})_*} & \pi_1(\check{\mathbb{P}}^2_{\mathbb{C}}\backslash\check{\Lambda}_{\mathcal{F}}) & \xrightarrow{\mu_{\mathcal{F}}} & \mathfrak{S}_d \\
\downarrow{\scriptstyle \imath_*} & & \downarrow{\scriptstyle \check{\imath}_*} & & \downarrow{\scriptstyle Id} \\
\pi_1(\mathbb{P}^2_{\mathbb{C}}\backslash\Delta_{\mathcal{F}}) & \xrightarrow{(\mathcal{G}^{\Delta}_{\mathcal{F}})_*} & \pi_1(\check{\mathbb{P}}^2_{\mathbb{C}}\backslash\check{\Delta}_{\mathcal{F}}) & \xrightarrow{\check{\mu}_{\mathcal{F}}} & \mathfrak{S}_d
\end{array},$$

es conmutativo, donde $\mathcal{G}^{\Delta}_{\mathcal{F}}$ denota la restricción de $\mathcal{G}_{\mathcal{F}}$ a $\mathbb{P}^2_{\mathbb{C}}\backslash\Delta_{\mathcal{F}}$, e $\imath:\mathbb{P}^2_{\mathbb{C}}\backslash\Lambda_{\mathcal{F}}\hookrightarrow\mathbb{P}^2_{\mathbb{C}}\backslash\Delta_{\mathcal{F}}$, $\check{\imath}:\check{\mathbb{P}}^2_{\mathbb{C}}\backslash\check{\Lambda}_{\mathcal{F}}\hookrightarrow\check{\mathbb{P}}^2_{\mathbb{C}}\backslash\check{\Delta}_{\mathcal{F}}$ son las inclusiones naturales. Por otro lado, por un teorema tipo Lefschetz [5], el morfismo $\imath_*$ es sobreyectivo. De la conmutatividad del diagrama anterior se sigue $\mathrm{Im}(\mathcal{G}^{\Lambda}_{\mathcal{F}})_*\subset\check{\imath}_*^{-1}(\mathrm{Im}(\mathcal{G}^{\Delta}_{\mathcal{F}})_*)$, y del hecho de que $\mathcal{G}^*_{\mathcal{F}}\mathrm{Leg}\,\mathcal{F}$ sea completamente descomponible resulta $\check{\imath}_*^{-1}(\mathrm{Im}(\mathcal{G}^{\Delta}_{\mathcal{F}})_*)\subset\check{\imath}_*^{-1}(\ker(\check{\mu}_{\mathcal{F}}))$. Las dos últimas propiedades permiten concluir la inclusión $\mathrm{Im}(\mathcal{G}^{\Lambda}_{\mathcal{F}})_*\subset\ker(\mu_{\mathcal{F}})$. □

Referencias

[1] E. Bertini; *Ricerche sulle trasformazioni univoche involutorie nel piano.* Annali di Mat., **8**: 244-286, 1877.

[2] A. Beltrán, M. Falla, D. Marín; *Flat 3-webs of degree one on the projective plane*, Annales de la Faculté des Sciencies de Toulouse. Vol. XXIII, n°4, 2014 pp. 779-796.

[3] A. Beltrán, M. Falla, D. Marín, M. Nicolau; *Foliations and rational maps inducing Galois coverings*, Preprint (2014).

[4] D. Cerveau and J. Deserti; *Feuilletages et transformations periódiques*, Experiments. Math. **19** (2010), 447-464.

[5] H. Hamm and L. Dung Trang; *Un théorème de Zariski du type Lefschetz*, Ann. Sci. Éc. Norm. Sup. **6** (1973), 317-366.

[6] S.L. Krushkal', B.N Apanasov and Gusevskii; *Kleinian Groups and Uniformization in Examples and Problems*, Translations of Mathematical Monographs **62**, AMS (1986).

[7] D. Marín and J.V. Pereira; *Rigid flat webs on the projective plane*, Asian Journal of Mathematics, **17** (2013), no. 1, 163-192.

[8] B. Maskit; *Kleinian groups*, Springer-Verlag (1988).

[9] M. Namba; *Branched coverings and algebraic functions*, Pitman Research Notes in Mathematics Series **161** (1987).

[10] J. V. Pereira; *Vector fields, invariant varieties and linear systems*, Ann. Inst. Fourier (Grenoble), **51** (2001), no. 5, 1385–1405.

Abstract

In this work we introduce the notion of Galois foliations on $\mathbb{P}^2_{\mathbb{C}}$, defined as those folations whose Gauss applications restricted to a Zariski open subset is a Galois covering. We also present some examples and a criterium for identifying such foliations.

Keywords: Foliations, webs.

Andrés Beltrán
Sección Matemáticas
Departamento de Ciencias
Pontificia Universidad Católica del Perú
Av. Universitaria 1801, Lima, Perú
abeltra@pucp.edu.pe

Maycol Falla
Departamento de Análise-IM
Universidade Federal Fluminense
Mário Santos Braga S/N- Niterói, 24.020-140 RJ, Brasil
maycolfl@gmail.com

David Marín
Departament de Matemátiques
Universitat Autònoma de Barcelona
E-08193 Bellaterra, Barcelona, España
davidmp@mat.uab.es

The Groebner basis of a polynomial system related to the Jacobian conjecture

Christian Valqui[1,2,3]***, Marco Solorzano***[1]

October, 2014

Abstract

We compute the Groebner basis of a system of polynomial equations related to the Jacobian conjecture using a recursive formula for the Catalan numbers.

MSC(2010): 14R15; 13F20, 11B99.

Keywords: Jacobian, Groebner basis, Catalan numbers.

[1] *Sección Matemáticas, Departamento de Ciencias, PUCP.*
[2] *Instituto de Matemáticas y Ciencias Afines (IMCA).*
[3] *Proyecto PUCP-DGI-2013-3036.*

1. Introduction

In this paper K is a characteristic zero field and $K[y]((x^{-1}))$ is the algebra of Laurent series in x^{-1} with coefficients in $K[y]$. In a recent article the following theorem was proved [3, Theorem 1.9].

Theorem 1.1. *The Jacobian conjecture in dimension two is false if and only if there exist*

- $P, Q \in K[x, y]$ *and* $C, F \in K[y]((x^{-1}))$,
- $n, m \in \mathbb{N}$ *such that* $n \nmid m$ *and* $m \nmid n$,
- $\nu_i \in K$ $(i = 0, \ldots, m + n - 2)$ *with* $\nu_0 = 1$,

such that

- C *has the form*

$$C = x + C_{-1}x^{-1} + C_{-2}x^{-2} + \cdots \qquad \text{with each } C_{-i} \in K[y],$$

- $gr(C) = 1$ *and* $gr(F) = 2 - n$, *where* gr *is the total degree,*
- $F_+ = x^{1-n}y$, *where* F_+ *is the term of maximal degree in* x *of* F,
- $C^n = P$ *and* $Q = \sum_{i=0}^{m+n-2} \nu_i C^{m-i} + F$.

Furthermore, under these conditions (P, Q) *is a counterexample to the Jacobian conjecture.* □

Motivated by this result, the authors consider the following slightly more general situation. Let D be a K-algebra (in Theorem 1.1 we take $D = K[y]$), n, m positive integers such that $n \nmid m$ and $n \nmid m$, $(\nu_i)_{0 \le i \le n+m-2}$ a family of elements in K with $\nu_0 = 1$, and $F_{1-n} \in D$ (in Theorem 1.1 we take $F_{1-n} = y$). A Laurent series in x^{-1} of the form

$$C = x + C_{-1}x^{-1} + C_{-2}x^{-2} + \cdots \qquad \text{with } C_{-i} \in D,$$

is **a solution of the system** $S(n, m, (\nu_i), F_{1-n})$ if there are $P, Q \in D[x]$ and $F \in D[[x^{-1}]]$, such that

$$F = F_{1-n}x^{1-n} + F_{-n}x^{-n} + F_{-1-n}x^{-1-n} + \cdots,$$
$$P = C^n, \qquad \text{and} \qquad Q = \sum_{i=0}^{m+n-2} \nu_i C^{m-i} + F.$$

For example, if $n = 2$, then

$$\begin{aligned} P(\mathbf{x}) = C^2 = {} & \mathbf{x}^2 + 2C_{-1} + 2C_{-2}\ \mathbf{x}^{-1} + (C_{-1}^2 + 2C_{-3})\ \mathbf{x}^{-2} \\ & + (2C_{-1}C_{-2} + 2C_{-4})\ \mathbf{x}^{-3} + (C_{-2}^2 + 2C_{-1}C_{-3} + 2C_{-5})\ \mathbf{x}^{-4} \\ & + (2C_{-2}C_{-3} + 2C_{-1}C_{-4} + 2C_{-6})\ \mathbf{x}^{-5} + \ldots, \end{aligned}$$

and the condition $C^2 \in K[x]$ translates into the following conditions on C_{-k}:

$$\begin{aligned} 0 &= (C^2)_{-1} = 2C_{-2}, \\ 0 &= (C^2)_{-2} = C_{-1}^2 + 2C_{-3}, \\ 0 &= (C^2)_{-3} = 2C_{-1}C_{-2} + 2C_{-4}, \\ 0 &= (C^2)_{-4} = C_{-2}^2 + 2C_{-1}C_{-3} + 2C_{-5}, \\ 0 &= (C^2)_{-5} = 2C_{-2}C_{-3} + 2C_{-1}C_{-4} + 2C_{-6}, \\ 0 &= (C^2)_{-6} = C_{-3}^2 + 2C_{-2}C_{-4} + 2C_{-1}C_{-5} + 2C_{-7}, \\ 0 &= (C^2)_{-7} = 2C_{-3}C_{-4} + 2C_{-2}C_{-5} + 2C_{-1}C_{-6} + 2C_{-8}, \\ 0 &= (C^2)_{-8} = C_{-4}^2 + 2C_{-3}C_{-5} + 2C_{-2}C_{-6} + 2C_{-1}C_{-7} + 2C_{-9}, \\ &\vdots \end{aligned}$$

In general, the condition $P(x) = C^n \in K[x]$ yields $(C^n)_{-k} = 0$, whereas $Q(x) = \sum_{i=0}^{m+n-2} \nu_i C^{m-i} + F \in K[x]$ handles us equations $\left(\sum_{i=0}^{m+n-2} \nu_i C^{m-i} + F\right)_{-k} = 0$, with $F_{-k} = 0$ for $k = 1, \ldots, n-2$.

It is easy to see (e.g. [3, Remark 1.13]) that the first $m + n - 2$ coefficients determine the others, i.e., the coefficients $C_{-1}, \ldots, C_{-m-n+2}$ determine univocally the coefficients C_{-k} for $k > m + n - 2$. Moreover,

the F_{-k} for $k > n-1$ depend only on F_{1-n} and C. Consequently, having a solution C to the system $S(n, m, (\nu_i), F_{1-n})$ is the same as having a solution $(C_{-1}, \ldots, C_{-m-n+2})$ to the system

$$\begin{aligned} E_k &= (C^n)_{-k} = 0, && \text{for } k = 1, \ldots, m-1, \\ E_{m-1+k} &= \left(\sum_{i=0}^{m+n-2} \nu_i C^{m-i} \right)_{-k} = 0, && \text{for } k = 1, \ldots, n-2, \\ E_{m+n-2} &= \left(\sum_{i=0}^{m+n-2} \nu_i C^{m-i} \right)_{1-n} + F_{1-n} = 0, && (1.1) \end{aligned}$$

with $m + n - 2$ equations $E_k = 0$ and $m + n - 2$ unknowns C_{-k}.

In order to understand the solution set of this system, it would be very helpful to find a Groebner basis for the ideal generated by the polynomials E_k in $D[C_{-1}, \ldots, C_{m+n-2}]$. In this paper we compute such a Groebner basis of (1.1) in a very particular case: we assume $n = 2$, $m = 2r + 1$ for some integer $r > 0$, and $\nu_i = 0$ for $i > 0$. Moreover, we consider $D = C[y]$ and $F_{1-n} = y$, as in Theorem 1.1.

2. Computation of a Groebner basis for I_{2r}

Assume $n = 2$, $m = 2r + 1$ for some integer $r > 0$, and $\nu_i = 0$ for $i > 0$. Set also $D = \mathbb{C}[y]$ and $F_{1-n} = y$.

Then the system (1.1) reads

$$E_i = \begin{cases} (C^2)_{-i}, & i = 1, \ldots, 2r \\ (C^{2r+1})_{-1} + y, & i = 2r + 1, \end{cases} \qquad (2.1)$$

where $(C^2)_{-i}$ denotes the coefficient of x^{-i} in the Laurent series C^2.

Explicitly, the polynomials E_i are given by

$$\begin{aligned}
E_1 &= 2C_{-2},\\
E_2 &= 2C_{-3}+(C_{-1})^2,\\
E_3 &= 2C_{-4}+2C_{-2}C_{-1},\\
E_4 &= 2C_{-5}+2C_{-3}C_{-1}+(C_{-2})^2,\\
E_5 &= 2C_{-6}+2C_{-2}C_{-3}+2C_{-4}C_{-1},\\
E_6 &= 2C_{-7}+2C_{-5}C_{-1}+2C_{-4}C_{-2}+(C_{-3})^2,\\
&\vdots\\
E_{2r-1} &= 2C_{-2r}+2C_{-2}C_{-2r+3}+2C_{-4}C_{-2r+5}+\cdots+2C_{-2r+4}C_{-3}+\\
&\quad 2C_{-2r+2}C_{-1},\\
E_{2r} &= 2C_{-2r-1}+2C_{-2r+1}C_{-1}+2C_{-2r+2}C_{-2}+\cdots+C_{-r}^2,\\
E_{2r+1} &= (C^{2r+1})_{-1}+y.
\end{aligned} \tag{2.2}$$

Each E_i is a polynomial in the ring $\mathbb{C}[C_{-1},C_{-2},\ldots,C_{-2r-1},y]$, and the $2r+1$ polynomials generate the ideal

$$I=\langle E_1,\ldots,E_{2r},E_{2r+1}\rangle.$$

Our goal is to find a Groebner basis for this I. However, in this section we will only compute a Groebner basis $(\widetilde{E}_1,\widetilde{E}_2,\ldots,\widetilde{E}_{2r-1},\widetilde{E}_{2r})$ for the ideal $I_{2r}=\langle E_1,E_2,\ldots,E_{2r-1},E_{2r}\rangle$.

Note that for $i=1\ldots,2r$ we have

$$E_i=2C_{-i-1}+\sum_{k=1}^{i-1}C_{-k}C_{k-i}. \tag{2.3}$$

We replace the odd numbered polynomials $E_1,E_3,E_5,E_7,\ldots,E_{2r-1}$

by new polynomials $\widetilde{E}_1, \widetilde{E}_3, \widetilde{E}_5, \widetilde{E}_7, \ldots, \widetilde{E}_{2r-1}$ defined by

$$\begin{aligned}
\widetilde{E}_1 &= C_{-2} = \frac{1}{2}E_1, \\
\widetilde{E}_3 &= C_{-4} = \frac{1}{2}E_3 - \widetilde{E}_1 C_{-1}, \\
\widetilde{E}_5 &= C_{-6} = \frac{1}{2}E_5 - \widetilde{E}_1 C_{-3} - \widetilde{E}_3 C_{-1}, \\
\widetilde{E}_7 &= C_{-8} = \frac{1}{2}E_7 - \widetilde{E}_1 C_{-5} - \widetilde{E}_3 C_{-3} - \widetilde{E}_5 C_{-1}, \\
\widetilde{E}_9 &= C_{-10} = \frac{1}{2}E_9 - \widetilde{E}_1 C_{-7} - \widetilde{E}_3 C_{-5} - \widetilde{E}_5 C_{-3} - \widetilde{E}_7 C_{-1}, \\
&\vdots \\
\widetilde{E}_{2r-1} &= C_{-2r} = \frac{1}{2}E_{2r-1} - \sum_{i=1}^{r-1} \widetilde{E}_{2i-1} C_{-2(r-i)+1}.
\end{aligned} \tag{2.4}$$

Remark 2.1. We have

$$\langle E_1, E_3, \ldots, E_{2r-1} \rangle = \langle \widetilde{E}_1, \widetilde{E}_3, \ldots, \widetilde{E}_{2r-1} \rangle.$$

In fact, if we define $\widetilde{I}_k^{odd} = \langle \widetilde{E}_1, \widetilde{E}_3, \ldots, \widetilde{E}_{2k-1} \rangle$, then (2.4) clearly implies

$$E_{2i+1} - 2\widetilde{E}_{2i+1} \in \widetilde{I}_i^{odd}, \tag{2.5}$$

and so we get $\langle E_1, E_3, \ldots, E_{2i+1} \rangle \subset \langle \widetilde{E}_1, \widetilde{E}_3, \ldots, \widetilde{E}_{2i+1} \rangle$ for $i = 0, 1, \ldots, r-1$. Using induction one sees that we also have $\langle \widetilde{E}_1, \widetilde{E}_3, \ldots, \widetilde{E}_{2r-1} \rangle \subset \langle E_1, E_3, \ldots, E_{2r-1} \rangle$, as desired.

The next proposition deals with $E_2, E_4, E_6, \ldots, E_{2r}$, the first r even numbered polynomials.

Proposition 2.2. *For all $j \in \mathbb{N}$ there exists λ_j such that for $\widetilde{E}_{2j} = C_{-2j-1} + \lambda_j C_{-1}^{j+1}$ we have*

$$C_{-2j-1} + \lambda_j C_{-1}^{j+1} - \frac{1}{2}E_{2j} \in \widetilde{I}_{2j-1} = \langle \widetilde{E}_1, \widetilde{E}_2, \ldots, \widetilde{E}_{2j-2}, \widetilde{E}_{2j-1} \rangle. \tag{2.6}$$

Moreover, if we set $\lambda_0 = -1$, then for $j > 0$, λ_j is given by

$$\lambda_j = \frac{1}{2}\left(\sum_{k=0}^{j-1} \lambda_k \lambda_{j-k-1}\right). \tag{2.7}$$

Proof. We proceed by induction on j. For $j = 0$ we set $\widetilde{E}_0 = 0$. Then we have

$$\widetilde{E}_0 \in \widetilde{I}_{2j-1} \quad \text{for all } j \geq 1, \quad \text{and} \quad \widetilde{E}_0 = C_{-1} + \lambda_0 C_{-1}. \tag{2.8}$$

For $j = 1$, with $\lambda_1 = \dfrac{1}{2}$ calculated by (2.7), we have

$$C_{-3} + \frac{1}{2}C_{-1}^2 - \frac{1}{2}E_2 = 0 \in \langle \widetilde{E}_1 \rangle,$$

as desired.

From (2.3) we have

$$\begin{aligned} E_{2j} = \ & 2C_{-2j-1} + \sum_{k=1}^{2j-1} C_{-k}C_{k-2j} \\ = \ & 2C_{-2j-1} + \sum_{k=0}^{j-1} C_{-2k-1}C_{2k+1-2j} + \sum_{k=1}^{j-1} C_{-2k}C_{2k-2j}, \end{aligned}$$

which clearly implies $\sum_{k=1}^{j-1} C_{-2k}C_{2k-2j} \in \widetilde{I}_{2j-1}$. Therefore we get

$$C_{-2j-1} - \frac{1}{2}E_{2j} \in -\frac{1}{2}\left(\sum_{k=0}^{j-1} C_{-2k-1}C_{2k+1-2j}\right) + \widetilde{I}_{2j-1}. \tag{2.9}$$

By the induction hypothesis and (2.8), for $0 \leq k \leq j-1$, there exist λ_k and λ_{j-k-1} such that

$$C_{-2k-1} = -\lambda_k C_{-1}^{k+1} + \widetilde{E}_{2k} \quad \text{and} \quad C_{2k+1-2j} = -\lambda_{j-k-1} C_{-1}^{j-k} + \widetilde{E}_{2(j-k-1)};$$

and hence

$$C_{-2k-1}C_{2k+1-2j} \in \lambda_k \lambda_{j-k-1} C_{-1}^{j+1} + \widetilde{I}_{2j-1}.$$

From (2.9) we obtain

$$C_{-2j-1} - \frac{1}{2}E_{2j} \in -\frac{1}{2}\left(\sum_{k=0}^{j-1} \lambda_k \lambda_{j-k-1}\right) C_{-1}^{j+1} + \widetilde{I}_{2j-1},$$

from which Relation (2.6) follows with $\lambda_j = \frac{1}{2}\left(\sum_{k=0}^{j-1} \lambda_k \lambda_{j-k-1}\right)$, as claimed. □

Corollary 2.3. *We have*

$$\langle E_1, E_2, \ldots, E_{2r}\rangle = \langle \widetilde{E}_1, \widetilde{E}_2, \ldots, \widetilde{E}_{2r}\rangle.$$

Proof. In fact, if we define $\widetilde{I}_k = \langle \widetilde{E}_1, \widetilde{E}_2, \ldots, \widetilde{E}_k\rangle$, then (2.5) and Proposition 2.2 imply

$$E_{k+1} - 2\widetilde{E}_{k+1} \in \widetilde{I}_k,$$

and so we get $\langle E_1, E_2, \ldots, E_{k+1}\rangle \subset \langle \widetilde{E}_1, \widetilde{E}_2, \ldots, \widetilde{E}_{k+1}\rangle$ for all k. Since we have $\langle E_1\rangle = \langle \widetilde{E}_1\rangle$, using induction one also obtains $\langle \widetilde{E}_1, \widetilde{E}_2, \ldots, \widetilde{E}_k\rangle \subset \langle E_1, E_2, \ldots, E_k\rangle$, as claimed. □

The bottom line of this corollary is that we can replace the system (2.2) with the following set of equations.

$$\begin{aligned} \widetilde{E}_1 &= C_{-2} = 0, \\ \widetilde{E}_3 &= C_{-4} = 0, \\ &\vdots \\ \widetilde{E}_{2r-1} &= C_{-2r} = 0, \end{aligned} \tag{2.10}$$

$$\begin{aligned}
\widetilde{E}_2 &= C_{-3} + \lambda_1 C_{-1}^2 = 0,\\
\widetilde{E}_4 &= C_{-5} + \lambda_2 C_{-1}^3 = 0,\\
&\vdots\\
\widetilde{E}_{2r} &= C_{-2r-1} + \lambda_r C_{-1}^{r+1} = 0,\\
\\
\widetilde{E}_{2r+1} &= (C^{2r+1})_{-1} + y = 0.
\end{aligned}$$

Proposition 2.4. *If we fix the lex order with $C_{-2r-1} > C_{-2r} > \cdots > C_{-3} > C_{-2} > C_{-1} > y$, then $G_{2r} = (\widetilde{E}_1, \widetilde{E}_2, \ldots, \widetilde{E}_{2r-1}, \widetilde{E}_{2r})$ is a Groebner basis of the ideal*

$$\widetilde{I}_{2r} = \langle \widetilde{E}_1, \widetilde{E}_2, \ldots, \widetilde{E}_{2r-1}, \widetilde{E}_{2r} \rangle$$

Proof. We first compute the S-polynomials of G_{2r}, and prove that they satisfy $\overline{S(\widetilde{E}_i, \widetilde{E}_j)}^{G_{2r}} = 0$ for $1 \leq i, j \leq 2r$.

Consider first the S-polynomial of an even-numbered polynomial and an odd-numbered polynomial, say $\widetilde{E}_{2s-1}$ and $\widetilde{E}_{2t}$, with $1 \leq s, t \leq r$. We have then

$$\begin{aligned}
S(\widetilde{E}_{2s-1}, \widetilde{E}_{2t}) &= C_{-2t-1}C_{-2s} - C_{-2s}(C_{-2t-1} + \lambda_t C_{-1}^{t+1})\\
&= -\lambda_t C_{-1}^{t+1} C_{-2s}\\
&= -\lambda_t C_{-1}^{t+1} \widetilde{E}_{2s-1},
\end{aligned}$$

and so $\overline{S(\widetilde{E}_{2s-1}, \widetilde{E}_{2t})}^{G_{2r}} = 0$, for all $1 \leq s, t \leq r$.

In case both i, j are odd, we take $\widetilde{E}_{2s-1}, \widetilde{E}_{2t-1}$, with $1 \leq s, t \leq r$. Then we have

$$S(\widetilde{E}_{2s-1}, \widetilde{E}_{2t-1}) = C_{-2t}C_{-2s} - C_{-2s}C_{-2t} = 0,$$

and trivially we get $\overline{S(\widetilde{E}_{2s-1}, \widetilde{E}_{2t-1})}^{G_{2r}} = 0$, for all $1 \leq s, t \leq r$.

In the last case, when i, j are even, consider $\widetilde{E}_{2s}, \widetilde{E}_{2t}$, with $1 \leq s, t \leq r$. Then we have

$$\begin{aligned} S(\widetilde{E}_{2s}, \widetilde{E}_{2t}) &= C_{-2t-1}(C_{-2s-1} + \lambda_s C_{-1}^{s+1}) - C_{-2s-1}(C_{-2t-1} + \lambda_t C_{-1}^{t+1}) \\ &= \lambda_s C_{-1}^{s+1} C_{-2t-1} - \lambda_t C_{-1}^{t+1} C_{-2s-1}. \end{aligned}$$

Now we divide $S(\widetilde{E}_{2s}, \widetilde{E}_{2t})$ by G_{2r}. If $C_{-2t-1} > C_{-2s-1}$, then the leading term is

$$lt(S(\widetilde{E}_{2s}, \widetilde{E}_{2t})) = \lambda_s C_{-1}^{s+1} C_{-2t-1},$$

and the first division step yields

$$S(\widetilde{E}_{2s}, \widetilde{E}_{2t}) = \lambda_s C_{-1}^{s+1} \widetilde{E}_{2t} + R_1,$$

with $R_1 = -\lambda_s \lambda_t C_{-1}^{s+t+2} - \lambda_t C_{-1}^{t+1} C_{-2s-1}$. By continuing the division algorithm we obtain

$$R_1 = -\lambda_t C_{-1}^{t+1} \widetilde{E}_{2s} + 0,$$

and hence $\overline{S(\widetilde{E}_{2s}, \widetilde{E}_{2t})}^{G_{2r}} = 0$ in this case. The case $C_{-2s-1} > C_{-2t-1}$ is similar, so we get $\overline{S(\widetilde{E}_{2t}, \widetilde{E}_{2s})}^{G_{2r}} = 0$ for $1 \leq s, t \leq r$. □

From Corollary 2.3 and Proposition 2.4 we are able conclude that $(\widetilde{E}_1, \widetilde{E}_2, \ldots, \widetilde{E}_{2r-1}, \widetilde{E}_{2r})$ is a Groebner basis for $\langle E_1, E_2, \ldots, E_{2r-1}, E_{2r} \rangle$.

3. A recursive formula for the Catalan numbers and a Groebner basis for the ideal

In this last section we will determine a Groebner basis for the ideal I given by the complete system (2.1). In order to achieve this we need to establish additional properties of the λ_j's which are closely related to the ubiquitous Catalan numbers.

Lemma 3.1. *For all $j \geq 0$ the equality*

$$c_j = (-1)^{j+1} 2^j \lambda_j \tag{3.1}$$

holds, where c_j are the Catalan numbers given by $c_j = \frac{1}{j+1}\binom{2j}{j}$.

Proof. The Catalan numbers are uniquely determined (see e.g. [4, p.117 (5.6)]) by $c_0 = 1$ and the recursive relation

$$c_r = \sum_{j=0}^{r-1} c_j c_{r-1-j}.$$

Set $d_j = (-1)^{j+1} 2^j \lambda_j$. Then $d_0 = 1$, since $\lambda_0 = -1$, and so equality (2.7) gives us

$$\begin{aligned} d_j =& (-1)^{j+1} 2^j \lambda_j \\ =& (-1)^{j+1} 2^j \frac{1}{2} \left(\sum_{k=0}^{j-1} \lambda_k \lambda_{j-k-1} \right) \\ =& \sum_{k=0}^{j-1} \left((-1)^{k+1} 2^k \lambda_k \right) \left((-1)^{j-k} 2^{j-1-k} \lambda_{j-k-1} \right) \\ =& \sum_{k=0}^{j-1} d_k d_{j-1-k}, \end{aligned}$$

and hence $d_j = c_j$ for all j, as desired. □

Now we prove a recursive formula for the Catalan numbers.

Proposition 3.2. *The Catalan numbers satisfy the following formula*

$$(2r+1)\frac{c_r}{2^{2r}} = \sum_{j=0}^{r} (-1)^j \binom{r}{j} \frac{c_j}{2^{2j}}. \tag{3.2}$$

Consequently, λ_r satisfies

$$(2r+1)(-1)^{r+1}\lambda_r = \sum_{j=0}^{r} \left(\begin{array}{c} r \\ j \end{array} \right) 2^{r-j}(-\lambda_j). \tag{3.3}$$

Proof. Replacing c_j in (3.2), and using (3.1) yields (3.3). Hence, it suffices to prove only (3.2). For that, we replace c_j by $\frac{1}{j+1}\binom{2j}{j}$ on the righthand side of (3.2) and use the equalities

$$\binom{-1/2}{j} = \frac{(-1)^j}{2^{2j}}\binom{2j}{j} \quad \text{and} \quad \binom{r+1/2}{r} = \frac{(2r+1)}{2^{2r}}\binom{2r}{r}.$$

Then we have

$$\begin{aligned}
\sum_{j=0}^{r}(-1)^j\binom{r}{j}\frac{c_j}{2^{2j}} &= \sum_{j=0}^{r}\frac{(-1)^j}{2^{2j}}\binom{2j}{j}\cdot\frac{1}{(j+1)}\binom{r}{j} \\
&= \sum_{j=0}^{r}\binom{-1/2}{j}\frac{1}{r+1}\binom{r+1}{j+1} \\
&= \frac{1}{(r+1)}\sum_{j=0}^{r}\binom{-1/2}{j}\cdot\binom{r+1}{r-j} \\
&= \frac{1}{(r+1)}\binom{r+1/2}{r} \\
&= \frac{1}{(r+1)}\frac{(2r+1)}{2^{2r}}\binom{2r}{r} \\
&= (2r+1)\frac{c_r}{2^{2r}}.
\end{aligned}$$

The second equality follows from $\frac{1}{j+1}\binom{r}{j} = \frac{1}{(r+1)}\binom{r+1}{j+1}$ and the fourth from $\binom{\alpha+\beta}{r} = \sum_{j=0}^{r}\binom{\alpha}{j}\binom{\beta}{r-j}$, relations valid for all $\alpha, \beta \in \mathbb{C}$. The last equality is known as the Chu–Vandermonde identity or Vandermonde convolution [1, p. 44, 13c′]. □

Proposition 3.3. *Let* $I_{2r} = \langle E_1, E_2, \ldots, E_{2r}\rangle$. *Then we have*

$$(C^{2r+1})_{-1} \in \mu_r C_{-1}^{r+1} + I_{2r},$$

for $\mu_r = \dfrac{2r+1}{(r+1)2^r}\binom{2r}{r}$.

Proof. By definition we have

$$(C^{2r+1})_{-1} = [(C^2)^r C]_{-1} = \sum_{j=-2}^{2r} [(C^2)^r]_j C_{-j-1},$$

since $C_{-j-1} = 0$ for $j < -2$ and $[(C^2)^r]_j = 0$ for $j > 2r$.

But we also have $[(C^2)^r]_j = \sum_{i_1+\cdots+i_r=j} (C^2)_{i_1} \dots (C^2)_{i_r}$. We claim that if $i_1 + \cdots + i_r = j$, then $i_k \geq -2r$ for $k = 1, \dots, r$. In fact, as $i_j \leq 2$, then so we have

$$i_1 + \cdots + i_{k-1} + i_{k+1} + \cdots + i_r \leq 2(r-1),$$

and $j = i_k + (i_1 + \cdots + i_{k-1} + i_{k+1} + \cdots + i_r) \leq 2(r-1) + i_k$ as well. Therefore we get $i_k \geq j - 2r + 2 \geq -2r$, since $j \geq -2$.

By definition we have $E_i = (C^2)_{-i}$ for $i = 1, \dots, 2r$. Consequently we obtain

$$(C^2)_{i_1} \dots (C^2)_{i_r} \in I_{2r}, \quad \text{if some } i_k \text{ is negative.}$$

It follows that

$$[(C^2)^r]_j \in \sum_{\substack{i_1+\cdots+i_r=j \\ i_k \geq 0}} (C^2)_{i_1} \dots (C^2)_{i_r} + I_{2r} = [(x^2 + 2C_{-1})^r]_j + I_{2r}$$

holds, since $C^2 = x^2 + 2C_{-1} + (C^2)_{-1}x^{-1} + (C^2)_{-2}x^{-2} + (C^2)_{-3}x^{-3} + \dots$. But we also have

$$(x^2 + 2C_{-1})^r = \sum_{k=0}^{r} \binom{r}{k} (2C_{-1})^{r-k} x^{2k},$$

and so

$$[(x^2 + 2C_{-1})^r]_j = \begin{cases} \binom{r}{k}(2C_{-1})^{r-k} & \text{if } j = 2k \\ 0, & \text{if } j = 2k+1. \end{cases}$$

We arrive at

$$(C^{2r+1})_{-1} \in \sum_{k=0}^{r} \binom{r}{k} (2C_{-1})^{r-k} C_{-2k-1} + I_{2r}.$$

Note that by Proposition 2.2 we have

$$C_{-2k-1} = \widetilde{E}_{2k} - \lambda_k C_{-1}^{k+1} \in -\lambda_k C_{-1}^{k+1} + I_{2r},$$

so we obtain

$$\begin{aligned}(C^{2r+1})_{-1} &\in \sum_{k=0}^{r} \binom{r}{k} (2C_{-1})^{r-k}(-\lambda_k C_{-1}^{k+1}) + I_{2r} \\ &= \left(\sum_{k=0}^{r} \binom{r}{k} 2^{r-k}(-\lambda_k)\right)(C_{-1})^{r+1} + I_{2r},\end{aligned}$$

and the formula for μ_r follows now from (3.1) and (3.3). □

Corollary 3.4. *For $\widetilde{E}_{2r+1} = \mu_r (C_{-1})^{r+1} + y$ we have*

$$\langle E_1, E_2, \ldots, E_{2r-1}, E_{2r}, E_{2r+1}\rangle = \langle \widetilde{E}_1, \widetilde{E}_2, \ldots, \widetilde{E}_{2r-1}, \widetilde{E}_{2r}, \widetilde{E}_{2r+1}\rangle.$$

Proof. By Proposition 3.3 we have $E_{2r+1} - \widetilde{E}_{2r+1} = (C^{2r+1})_{-1} - \mu_r C_{-1}^{r+1} \in I_{2r}$. The result follows now from Corollary 2.3. □

Now we can state our main result.

Theorem 3.5. *If we fix the lex order with $C_{-2r-1} > C_{-2r} > \cdots > C_{-3} > C_{-2} > C_{-1} > y$, then $G_{2r+1} = (\widetilde{E}_1, \widetilde{E}_2, \ldots, \widetilde{E}_{2r}, \widetilde{E}_{2r+1})$ is a Groebner basis for the ideal*

$$I = \langle E_1, E_2, \ldots, E_{2r-1}, E_{2r}, E_{2r+1}\rangle.$$

Proof. By Corollary 3.4 it suffices to prove that the division of the S-polynomials $S(\widetilde{E}_i, \widetilde{E}_j)$ by G_{2r+1} is zero. If $i, j \leq 2r$, then the division algorithm yields the same quotients and remainders as in Proposition 2.4, since the remainders become zero before one has to divide by $\widetilde{E}_{2r+1}$. Note that $lt(\widetilde{E}_{2r+1}) = \mu_r (C_{-1})^{r+1}$, since $\mu_r \neq 0$. It remains to divide the S-polynomials $S(\widetilde{E}_i, \widetilde{E}_{2r+1})$ by G_{2r+1}. We first consider the case $i = 2t-1$ for some $t = 1, \ldots, r$. We get

$$\begin{aligned}S(\widetilde{E}_{2t-1}, \widetilde{E}_{2r+1}) &= \frac{C_{-2t}C_{-1}^{r+1}}{C_{-2t}}(C_{-2t}) - \frac{C_{-2t}C_{-1}^{r+1}}{\mu_r C_{-1}^{r+1}}(\mu_r C_{-1}^{r+1} + y) \\ &= -\frac{1}{\mu_r} y C_{-2t},\end{aligned}$$

for all $t = 1, \ldots, r$. The first division step yields $S(\widetilde{E}_{2t-1}, \widetilde{E}_{2r+1}) = -\frac{1}{\mu_r} y \widetilde{E}_{2t-1}$, hence we obtain $\overline{S(\widetilde{E}_{2t-1}, \widetilde{E}_{2r+1})}^{G_{2r+1}} = 0$, for all $t = 1, \ldots, r$.

Now for the S-polynomials of $\widetilde{E}_{2t}$ and $\widetilde{E}_{2r+1}$, for some $t = 1, \ldots, r$, we have

$$\begin{aligned} S(\widetilde{E}_{2t}, \widetilde{E}_{2r+1}) &= \frac{C_{-2t-1}C_{-1}^{r+1}}{C_{-2t-1}}(C_{-2t-1} + \lambda_t C_{-1}^{t+1}) - \\ &\qquad \frac{C_{-2t-1}C_{-1}^{r+1}}{\mu_r C_{-1}^{r+1}}(\mu_r C_{-1}^{r+1} + y) \\ &= \lambda_t C_{-1}^{r+t+2} - \frac{1}{\mu_r} C_{-2t-1} y. \end{aligned}$$

with leading term

$$lt(S(\widetilde{E}_{2t}, \widetilde{E}_{2r+1})) = -\frac{1}{\mu_r} C_{-2t-1} y.$$

We divide $S(\widetilde{E}_{2t}, \widetilde{E}_{2r+1})$ by G_{2r+1}, and the first division step gives us

$$S(\widetilde{E}_{2t}, \widetilde{E}_{2r+1}) = -\frac{1}{\mu_r} y \widetilde{E}_{2t} + R_1$$

with $R_1 = \lambda_t C_{-1}^{r+t+2} + \frac{\lambda_t}{\mu_r} y C_{-1}^{t+1}$. Finally we take note of the equality $R_1 = \frac{\lambda_t}{\mu_r} C_{-1}^{t+1} \widetilde{E}_{2r+1}$, in order to obtain $\overline{S(\widetilde{E}_{2t}, \widetilde{E}_{2r+1})}^{G_{2r+1}} = 0$, for all $t = 1, \ldots, r$. This concludes the proof. □

In brief, we give the Groebner basis $G_{2r+1} = (\widetilde{E}_1, \widetilde{E}_2, \ldots, \widetilde{E}_{2r}, \widetilde{E}_{2r+1})$ of I explicitly as

$$\begin{aligned}
\widetilde{E}_1 &= C_{-2}, \\
\widetilde{E}_3 &= C_{-4}, \\
&\vdots \\
\widetilde{E}_{2r-1} &= C_{-2r}, \\
\\
\widetilde{E}_2 &= C_{-3} + \lambda_1 C_{-1}^2, \\
\widetilde{E}_4 &= C_{-5} + \lambda_2 C_{-1}^3, \\
&\vdots \\
\widetilde{E}_{2r} &= C_{-2r-1} + \lambda_r C_{-1}^{r+1}, \\
\\
\widetilde{E}_{2r+1} &= \mu_r (C_{-1})^{r+1} + y.
\end{aligned}$$

with

$$\mu_r = \frac{2r+1}{(r+1)2^r}\binom{2r}{r} \quad \text{and} \quad \lambda_j = \frac{(-1)^{j+1}}{(j+1)2^j}\binom{2j}{j}.$$

Acknowledgments

We wish to thank Jonathan Farfán, who provided the proof of Formula (3.3) (and hence, implicitly, of Formula (3.1)).

References

[1] L. Comtet; *Advanced combinatorics*, D. Reidel Publishing Co., Dordrecht, 1974.

[2] D. Cox, J. Little, D. O'Shea; *Ideals, varieties, and algorithms*, Undergraduate Texts in Mathematics, 3, Springer, New York, 2007.

[3] J.A. Guccione, J.J. Guccione, C. Valqui; *A system of polynomial equations related to the Jacobian conjecture*, arXiv:1406.0886v1 [math.AG] (3 June 2014).

[4] T. Koshy; *Catalan numbers with applications*, Oxford University Press, Oxford, 2009.

Resumen

En este artículo calculamos la base de Groebner de un sistema polinomial de ecuaciones relacionada con la conjetura del jacobiano utilizando una formula recursiva para los números de Catalan.

Palabras clave: Jacobiano, bases de Groebner, números de Catalan

Christian Valqui
Sección Matemáticas
Departamento de Ciencias
Pontificia Universidad Católica del Perú
Av. Universitaria 1801, San Miguel, Lima 32, Perú

Instituto de Matemática y Ciencias Afines (IMCA)
Calle Los Biólogos 245, Urb. San Cesar, La Molina, Lima 12, Perú
cvalqui@pucp.edu.pe

Marco Solorzano
Sección Matemáticas
Departamento de Ciencias
Pontificia Universidad Católica del Perú
Av. Universitaria 1801, San Miguel, Lima 32, Perú
marco.solorzano@pucp.pe

Reducción del grado en aplicaciones de Keller

***P. Fernandez*[1,2], *R. Rabanal*[1,2,3]**

Octubre, 2014

Resumen

A las aplicaciones polinomiales con el determinante de su matriz jacobiana igual a 1 se las llama aplicaciones de Keller. Según la conjetura jacobiana de Keller, cada aplicación de Keller es inyectiva. Tal conjetura es verdadera para las aplicaciones polinomiales de grado menor o igual a dos. En el presente trabajo también se muestra que el caso general se reduce a estudiar la inyectividad de aplicaciones de la forma $z \mapsto z + H(z)$, donde las componentes no nulas de H son polinomios homogeneos de grado tres y cada matriz Jacobiana $DH(z)$ es nilpotente.

MSC(2010): 14R15; 13F20.

Palabras Clave: Aplicacion Keller, anillo de polinomios, automorfismo.

[1] *Sección Matemáticas, Departamento de Ciencias, PUCP.*
[2] *Proyecto PUCP-DGI 2010-0058.*
[3] *ICTP-ITALY (220 (Maths) RR/ab).*

1. Introducción

Este artículo pretende divulgar algunos resultados clásicos relacionados con la inyectividad global de los difeomorfismos locales [18, 20, 5, 3, 15, 16, 9, 19, 4]. Sólo describimos los difeomorfismos locales inducidos por aplicaciones polinomiales con coeficientes complejos; en particular, aquellos que son obtenidos por las **aplicaciones de Keller**. Esto significa que la aplicación es polinomial y el determinante de cada matriz jacobiana es igual a uno [12]. Un ejemplo inicial de este tipo de aplicaciones es originado por una polinomial inyectiva que preserva el área; esto porque su inversa también es polinomial, y será Keller por una aplicación directa de la regla de la cadena (ver § **3.2**). Recíprocamente, no es difícil comprobar que cada aplicación de Keller lineal es una biyección, y lo mismo sucederá si tal aplicación estuviera formada por polinomios de grado menor o igual a dos (ver § **3.3**). Sin embargo, aún no se conoce una respuesta general a la **conjetura jacobiana de Keller**, según la cual cada aplicación de Keller es injectiva [12, 21, 10, 2, 23] (ver también [9, 3, 5, 17]).

Dar una solución a tal conjetura consiste en demostrar que cada aplicación de Keller es inyectiva, o en su defecto encontrar explícitamente una aplicación de Keller no inyectiva. Para hacer esto es necesario examinar atentamente las aplicaciones polinomiales inyectivas. A este estudio se ha dedicado la sección 2, donde mostramos que las polinomiales inyectivas también son dominantes, porque son sobreyectivas (ver § **2.4**) y, consecuentemente, su inversa no sólo es birracional sino también polinomial. Finalmente, en la sección 3 mostramos que por una sucesión finita de cambios de variables, estudiar la conjetura jacobiana de Keller se reduce a trabajar con aplicaciones polinomiales de grado menor o igual a tres. Precisamente, en § **4.8** se demuestra que *para resolver la conjetura jacobiana de Keller bastará estudiar la inyectividad de todas la aplicaciones de la forma* $z \mapsto z + H(z)$, *donde las componentes no nulas de* H *son polinomios homogéneos de grado tres y cada derivada* $DH(z)$ *es nilpotente [2].*

2. Aplicaciones polinomiales e inyectividad

Describimos algunos resultados básicos acerca de las aplicaciones polinomiales inyectivas. Concluimos con el apartado § **2.6**, mostrando que cada polinomial inyectiva es una biyección con inversa polinomial.

2.1. Sea $\mathbb{C}[X_1,\dots,X_n]=\mathbb{C}[X]$ el anillo de los polinomios en n variables con coeficientes en el cuerpo de los números complejos. Cada subconjunto ordenado $(F_1,\dots,F_n)$ de $\mathbb{C}[X_1,\dots,X_n]$ determina con exactitud una aplicación del espacio vectorial $\mathbb{C}^n=\{z=(z_1,\dots,z_n):z_1,\dots,z_n\in\mathbb{C}\}$ en sí mismo. Esta aplicación, definida por la regla $z\mapsto F(z)=\big(F_1(z),\dots,F_n(z)\big)$, se denomina **aplicación polinomial** y se denota por $F=(F_1,\dots,F_n)$. Como de costumbre, la regla $z\mapsto F_i(z)$ define una de las funciones **coordenadas** de F, que se identifica con el polinomio F_i.

2.2. Cada polinomial F está asociada a un único conjunto ordenado $(F_1,\dots,F_n)$, el cual induce el anillo $\mathbb{C}[F_1,\dots,F_n]$. Si $\mathbb{C}[X_1,\dots,X_n]$ y $\mathbb{C}[F_1,\dots,F_n]$ son iguales, se dice que la aplicación es **algebraicamente invertible**. En este caso, existe algún polinomio $G_i(F_1,\dots,F_n)$ con $G_i(F_1,\dots,F_n)=X_i$; por eso $G=(G_1,\dots,G_n)$ es una inversa a izquierda de F. Por otro lado, si la polinomial $H=(H_1,\dots,H_n)$ es una inversa a izquierda de F, cada variable X_i es $H_i(F_1,\dots,F_n)$ y, así, está en $\mathbb{C}[F_1,\dots,F_n]$. En conclusión, *algebraicamente invertible equivale a la existencia de alguna polinomial como inversa a izquierda.*

2.3. A partir de $\mathbb{C}[X_1,\dots,X_n]$ se define $\mathbb{C}(X_1,\dots,X_n)$, su cuerpo de fracciones. Análogamente, cada polinomial $F=(F_1,\dots,F_n)$ origina el cuerpo $\mathbb{C}(F_1,\dots,F_n)$. Si ambos cuerpos de fracciones son iguales, la polinomial se denomina **birracional**. En este caso, cada X_i se escribe como un cociente $\dfrac{g_i(F_1,\dots,F_n)}{h_i(F_1,\dots,F_n)}$ y, así, la aplicación $\psi:z\mapsto\big(\dfrac{g_1(z)}{h_1(z)},\dots,\dfrac{g_n(z)}{h_n(z)}\big)$ es una inversa a izquierda de F. Recíprocamente, cuando F tiene alguna aplicación racional como inversa a izquierda, $\mathbb{C}(X_1,\dots,X_n)$ es un subcuerpo de $\mathbb{C}(F_1,\dots,F_n)$ y por eso

son iguales. En el acápite § **2.5** se verá que la aplicación $\mathbb{C}[F_1,\ldots,F_n] \ni \phi \mapsto \phi \circ F \in \mathbb{C}[X_1,\ldots,X_n]$ induce $\mathbb{C}(F_1,\ldots,F_n) \hookrightarrow \mathbb{C}(X_1,\ldots,X_n)$. Por tanto, *ser birracional equivale a la existencia de alguna aplicación racional como inversa a izquierda.*

2.4. La inyectividad de $F = (F_1,\ldots,F_n)$ se describe vía los ceros de algunos ideales de $\mathbb{C}[X_1,\ldots,X_n,Y_1,\ldots,Y_n] = \mathbb{C}[X,Y]$. Ello es posible porque F es inyectiva si y solo si cada solución $(z,w) \in \mathbb{C}^n \times \mathbb{C}^n$ del sistema $F_i(X) - F_i(Y) = 0, j = 1,\ldots,n$, también es solución del sistema $X_i - Y_i = 0, j = 1,\ldots,n$. Esto significa que los ceros del ideal $I = \big(F_1(X) - F_1(Y);\cdots;F_n(X) - F_n(Y)\big)$ pertenecen al conjunto de ceros $V(J) = \{(z,w) \in \mathbb{C}^n \times \mathbb{C}^n : f(z,w) = 0, f \in J\}$, donde J es el ideal generado por $X_1 - Y_1, X_2 - Y_2, \ldots, X_n - Y_n$. Por el teorema de los ceros de Hilbert [13, Nullstellensatz], la inclusión $V(I) \subset V(J)$ implica que para algún $m \in \mathbb{N}$ se tiene $(X_i - Y_i)^m \in I$, es decir existen polinomios $h_{i,k} \in \mathbb{C}[X,Y]$ sujetos a

$$(X_i - Y_i)^m = \sum_{k=1}^{n} h_{i,k}(X,Y)[F_k(X) - F_k(Y)]. \tag{2.1}$$

Por medio de esta identidad se prueba que *cada polinomial inyectiva es sobreyectiva* [1]. Como en [11, 21, 6], se procede por contradicción y para alguna aplicación inyectiva $(F_1,\ldots,F_n)$ se elige un $c = (c_1,\ldots,c_n) \in \mathbb{C}^n$ de modo que el sistema $F_i(X) - c_i, i = 1,\ldots,n$, no tenga solución en $\mathbb{C}^n$; en otras palabras, la variedad $V(F_1(X) - c_1;\cdots;F_n(X) - c_n)$ resulta vacía y por lo tanto igual a $V(1)$. En consecuencia existirián polinomios $h_i \in \mathbb{C}[X,Y]$ que cumplirán

$$1 = \sum_{i=1}^{n} h_i(X,Y)[F_i(X) - c_i]. \tag{2.2}$$

Con esto, se considera el conjunto $\{d_1,\ldots,d_t\}$ formado por la unión de $\{c_1,\ldots,c_n\}$ con los coeficientes de todos los polinomios $F_i(X)$, $h_{i,k}(X,Y)$, $h_i(X,Y)$. A partir de ahí se construye el subanillo $\mathbb{Z}[d_1,\ldots,d_t]$ de $\mathbb{C}$, una $\mathbb{Q}-$álgebra. Por el lema de normalización de

Noether, como aparece en Nagata [14, 22] (ver apendice), existe un ideal maximal $\mathfrak{m}$ del anillo $\mathbb{Z}[d_1,\ldots,d_t]$ de modo que el cociente $K = \mathbb{Z}[d_1,\ldots,d_t]/\mathfrak{m}$ es un cuerpo finito. Por otro lado, por cada subconjunto ordenado $(\eta_1,\ldots,\eta_n)$ de $\mathbb{Z}[d_1,\ldots,d_t]$ se cumple que la imagen $F_i(\eta_1,\ldots,\eta_n)$ también está en $\mathbb{Z}[d_1,\ldots,d_t]$, pues los coeficientes de F_i están en $\{d_1,\ldots,d_t\}$. Por eso $\phi : (\eta_1,\ldots,\eta_n) \mapsto \big(F_1(\eta_1,\ldots,\eta_n),\ldots,F_n(\eta_1,\ldots,\eta_n)\big)$ determina una polinomial en el producto $(\mathbb{Z}[d_1,\ldots,d_t])^n$, y su cociente módulo $\mathfrak{m}$ induce una aplicación ϕ_0 por medio del diagrama

$$\begin{array}{ccc} (\mathbb{Z}[d_1,\ldots,d_t])^n & \xrightarrow{\phi} & (\mathbb{Z}[d_1,\ldots,d_t])^n \\ \alpha\downarrow & & \downarrow\alpha \\ K^n & \xrightarrow{\phi_0} & K^n, \end{array}$$

donde $\alpha(\eta_1\ldots,\eta_n) = (\overline{\eta}_1\ldots,\overline{\eta}_n)$, con $\overline{\eta}_i$ la clase de η_i módulo $\mathfrak{m}$. Más aún, para ϕ_0 se cumplen dos propiedades. Primero, por (2.1) y la inyectividad de $F = (F_1,\ldots,F_n)$ en $\mathbb{C}^n$, la aplicación ϕ_0 es inyectiva. Segundo, a partir de (2.2) y la elección de $c = (c_1,\ldots,c_n)$, tal ϕ_0 no es sobreyectiva, pues $\alpha(c) \in K^n \setminus \phi_0(K^n)$. Esta contradicción entre apliaciones inyectivas de conjuntos finitos muestra que F es sobreyectiva. Por tanto, *cada aplicación polinomial inyectiva es sobreyectiva.*

2.5. Cada $F = (F_1,\ldots,F_n)$ inyectiva en $\mathbb{C}^n$ también es sobreyectiva. En particular, la imagen $F(\mathbb{C}^n)$ es densa en $\mathbb{C}^n$, lo que significa que F es una aplicación **dominante**. En consecuencia, cuando φ y ψ están en $\mathbb{C}[F_1,\ldots,F_n]$, la condición $\varphi\circ F = \psi\circ F$ en $\mathbb{C}[X_1,\ldots,X_n]$ implica $\varphi = \psi$ en $\mathbb{C}[F_1,\ldots,F_n]$. En otras palabras, el morfismo de anillos F^* definido por $\mathbb{C}[F_1,\ldots,F_n] \ni \phi \mapsto \phi\circ F \in \mathbb{C}[X_1,\ldots,X_n]$ es inyectivo. Más aún, decir que F es dominante equivale a la inyectividad de F^*, el cual induce un monomorfismo entre los cuerpos de fracciones $\mathbb{C}(F_1,\ldots,F_n)$ y $\mathbb{C}(X_1,\ldots,X_n)$, el mismo que será sobreyectivo, pues el dominio y la imagen de F tienen igual dimensión. En efecto, el monomorfismo F^* induce una extensión finita de cuerpos $\mathbb{C}(F_1,\ldots,F_n) \hookrightarrow \mathbb{C}(X_1,\ldots,X_n)$ en el sentido que $\mathbb{C}(X_1,\ldots,X_n)$, es un espacio vectorial sobre

$\mathbb{C}(F_1, \ldots, F_n)$ y su dimensión $d(F)$ es finita. Esta dimensión $d(F)$ es el **grado geométrico** de F, y obviamente cuando $d(F) = 1$ los cuerpos de fracciones son iguales. Más aún, $d(F)$ describe las fibras $F^{-1}(w) = \{v : F(v) = w\}$ de F. Es decir, *existe un conjunto denso $U \subset \mathbb{C}^n$ tal que para todo $w \in U$ el grado $d(F)$ es igual a $\#(F^{-1}(w))$, el número de elementos de $F^{-1}(w)$.* En conclusión, si $F = (F_1, \ldots, F_n)$ es inyectiva, entonces $d(F) = \#(F^{-1}(w)) = 1$ y por consiguiente ambos cuerpos de fracciones, $\mathbb{C}(X_1, \ldots, X_n)$ y $\mathbb{C}(F_1, \ldots, F_n)$, coinciden y satisfacen la definición de § **2.3**. *Por tanto, cada polinomial inyectiva es una aplicación birracional.*

2.6. Cuando $F = (F_1, \ldots, F_n)$ es inyectiva, admite una inversa a la izquierda y, por § **2.4**, es una biyección. Para demostrar que su inversa es polinomial, se usa la sobreyectividad y § **2.5**, el cual asegura la existencia de polinomios que tienen a $1 \in \mathbb{C}$ como único divisor común, cuyo cociente satisface $\dfrac{g_i(F_1, \ldots, F_n)}{h_i(F_1, \ldots, F_n)} = X_i$. A partir de esto, se obtiene que F es algebraicamente invertible cuando $h_i \in \mathbb{C}^*$ para todo i. Si por el contrario, algún polinomio h_j no perteneciera a $\mathbb{C}^*$, entonces la variedad $V(h_j) = \{z \in \mathbb{C}^n : h_j(z) = 0\}$ sería un subconjunto no vacío de $V(g_j)$, porque cumplen $g_j(z) = w_j h_j(z)$ cuando $z = F(w_1, \ldots, w_n)$. Por el teorema de los ceros de Hilbert [13, 22], la inclusión $V(h_j) \subset V(g_j)$ implica que g_j^m, para algún $m \in \mathbb{N}$, es multiplo de h_j. Esta contradicción entre los divisores comunes de g_j y h_j muestra $h_j \in \mathbb{C}^*$ (vea [6, 12]). Por tanto, *cada polinomial inyectiva es una biyección algebraicamente invertible y su inversa también es polinomial.*

3. Aplicaciones Keller de grado pequeño

En esta sección probamos un caso particular de la conjetura jacobiana de Keller. Específicamente, demostramos la inyectividad de las aplicaciones de Keller cuyas funciones coordenadas son polinomios de grado menor o igual a dos (ver § **3.3**). Este resultado fue reportado por Wang [23], pero la demostración que presentamos es adoptada de [2].

3.1. Cada aplicación polinomial $F = (F_1, \ldots, F_n)$ es holomorfa en $\mathbb{C}^n$. Gracias a esto siempre existen las derivadas parciales $\frac{\partial F_i}{\partial x_j}$ y queda definida la **matriz jacobiana**

$$DF(z) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(z) & \cdots & \frac{\partial F_1}{\partial x_n}(z) \\ \vdots & & \vdots \\ \frac{\partial F_n}{\partial x_1}(z) & \cdots & \frac{\partial F_n}{\partial x_n}(z) \end{pmatrix}.$$

Más aún, su determinante $det(DF(z))$ es un polinomio con algún cero de no ser constante, ya que $\mathbb{C}$ es algebraicamente cerrado. Por tanto, *las afirmaciones* $det(DF(z)) \neq 0$, *para todo* $z \in \mathbb{C}^n$ *y existe* $c \in \mathbb{C}^* = \mathbb{C}\backslash\{0\}$ *tal que* $det(DF(z)) = c$, *para todo* $z \in \mathbb{C}^n$ *son equivalentes.*

3.2. El apartado § **2.6** hizo patente que las aplicaciones polinomiales inyectivas son biyecciones con inversa polinomial. La regla de la cadena demuestra que cada matriz jacobiana tiene inversa. De esta manera, *las aplicaciones polinomiales inyectivas que preservan el área son Keller.*

3.3. Cada aplicación polinomial con polinomios de grado uno o cero es una transformación lineal y su matriz asociada en la base canónica es la matriz jacobiana. Luego, cada aplicación de Keller lineal es una biyección. Lo mismo sucede si en la aplicación de Keller $F = (F_1, \ldots, F_n)$ los polinomios F_i son de grado menor o igual a dos, porque en este caso la condición $F(a) = F(b)$ implica dos cosas. Primero, la aplicación $G(z) = F(z + a) - F(b)$ satisface $G(b - a) = G(0) = 0$ y por eso se escribe $G = G_{(1)} + G_{(2)}$, donde cada $G_{(j)}$ es una aplicación polinomial formada por polinomios de grado $j \in \{1, 2\}$. Segundo, al jugar con cierta evaluación y derivada de $t \to G(t(b - a)) = tG_{(1)}(b - a) + t^2G_{(2)}(b - a)$ se obtiene

$$\begin{aligned} 0 = G(b - a) &= \left\{G_{(1)}(b - a) + 2tG_{(2)}(b - a)\right\}_{t=\frac{1}{2}} \\ &= \left\{DF(t(b - a)) \cdot (b - a)\right\}_{t=\frac{1}{2}}. \end{aligned}$$

Como la matriz $DF(\frac{1}{2}(b-a))$ es invertible concluimos que, $F(a) = F(b)$ implica $a = b$. Por tanto, *cada aplicación de Keller es inyectiva, si sus componentes son de grado menor o igual a dos* (ver también [23, 15]).

4. Reducción del grado

El grado de una aplicación polinomial $F = (F_1, \ldots, F_n)$ es el mayor grado de las componentes F_i, es decir $deg(F) = \max_i deg(F_i)$, donde $deg(F_i)$ es el grado de las componentes en el sentido usual.

4.1. Para estudiar la inyectividad de F uno se toma la libertad de reemplazar F por $H \circ F \circ \widetilde{H}$ donde H y $\widetilde{H}$ son isomorfismos afines. En particular, se puede cambiar F por $T \circ F$ con T la traslación $T(z) = z - F(0)$ y trabajar con aplicaciones que envían el cero en el cero. Del mismo modo, si $DF(0)$ tiene inversa, al sustituir F por $(DF(0))^{-1} \circ F$ se obtiene una nueva aplicación cuyas componentes se escriben de la forma $X_i +$ términos de grado al menos dos. Por lo tanto, para estudiar la inyectividad se puede suponer que *la aplicación F satisface $F(0) = 0$ y que su derivada $DF(0)$ es la matriz identidad I_n.*

Observación 4.2. Gracias a la libertad concedida por el acápite § 4.1, en el estudio de la conjetura jacobiana de Keller muchos autores trabajan con polinomiales para las cuales el determinate de su matriz jacobiana es constante. Por otro lado, ya es conocido que la equivalencia anunciada en § **3.1** no es verdadera para aplicaciones reales $\mathbb{R}^n \mapsto \mathbb{R}^n$. Más aún, Pinchuck describe en [17] una polinomial real que no es inyectiva, aún cuando el determinante de su matriz jacobiana nunca se anula.

4.3. Sean P y Q dos polinomios en $\mathbb{C}[X_1, \ldots, X_n]$. La inyectividad de la aplicación de Keller $F = (F_1, \ldots, F_n)$ es equivalente a la inyectividad de la nueva aplicacion de Keller $F^{[2]}$, definida en $\mathbb{C}^n \times \mathbb{C}^2 = \{(z, w) : z \in \mathbb{C}^n, w \in \mathbb{C}^2\}$ por la regla $(z, w) \mapsto (F(z), w_1 - P(z), w_2 - Q(z))$ como es fácil observar. Tal equivalencia persiste, si se considera la composición $E_1 \circ F^{[2]} = (\widetilde{F}_1, \ldots, \widetilde{F}_{n+2})$ con $E_1(z, w) = (z_1 - w_1 w_2, z_2, \ldots, z_n, w_1, w_2)$, que también es Keller. Por esta razón, cuando la componente $F_1 \in \mathbb{C}[X_1, \ldots, X_n]$ admite al producto PQ entre sus monomios de grado $d \geq 4$, en la componente modificada $\widetilde{F}_1(z, w) = [F_1(z) - P(z)Q(z)] - w_1 w_2 + w_1 Q(z) + w_2 P(z)$ ya no aparece PQ y los tres términos $w_1 w_2$, $w_1 Q(z)$ y $w_2 P(z)$ son de grado a lo mucho $d - 1$. Es decir, para

estudiar la inyectividad, aumentado inductivamente las variables de ser necesario, será suficiente considerar aplicaciones polinomiales cuya primera componente tiene grado menor o igual a tres. Todo esto se puede repetir por cada componente. Por tanto, *la inyectividad de la aplicación de Keller* $F = (F_1, \ldots, F_n)$, *con grado al menos* 4, *es equivalente a la inyectividad de alguna aplicación de Keller* $G = (G_1, \ldots, G_{n+\tilde{m}})$ *con* $\tilde{m} \geq 2$ *y* $deg(G) \leq 3$ (ver [2]).

Observación 4.4. Si algún $X_j^{n_j}$ con $n_j \geq 2$ dividiera a un monomio $M = M(z, w)$ de la primera componente de $G = (G_1, \ldots, G_{n+\tilde{m}})$, la técnica de § **4.3** se podría repetir factorizando $M = PQ$ de modo que X_j divida a cada uno de los polinomios, a P y a Q. Éste proceso permite encontrar $(\widetilde{G}_1, \ldots, \widetilde{G}_{n+\tilde{m}+2})$, donde $\widetilde{G}_1(z, w, \widetilde{w}) = [G_1(z, w) - P(z, w)Q(z, w)] - \widetilde{w}_1\widetilde{w}_2 + \widetilde{w}_1 Q(z, w) + \widetilde{w}_2 P(z, w)$, y así $M = PQ$ no aparece como un monomio y, además, la potencia de X_j esta vez será $n_j - 1$. Por lo tanto, podemos concluir que las componentes de la aplicación del parágrafo anterior son lineales en cada variable.

4.5. Cuando la aplicación $F = (F_1, \ldots, F_n)$ de § **4.3** satisface las condiciones de § **4.1**, la aplicación $G = (G_1, \ldots, G_{n+\tilde{m}})$ envía el origen en el origen y su matriz jacobiana $DG(0)$ es la identidad $I_{n+\tilde{m}}$. Por tanto, *para investigar la inyectividad de una aplicación de Keller es suficiente estudiar aquellas de la forma*

$$z \mapsto z + F_{(2)}(z) + F_{(3)}(z),$$

donde cada componente de $F_{(i)}$ *o es cero o es un polinomio homogéneo de grado* $i \in \{2, 3\}$. (Por la observación 4.4, inclusive se puede conseguir que las componentes homogéneas estén formadas por monomios que son lineales en cada variable).

4.6. A partir de $F(z) = z + F_{(2)}(z) + F_{(3)}(z)$ se define en $\mathbb{C}^p \times \mathbb{C}^p = \{(z, \tilde{z}) : z, \tilde{z} \in \mathbb{C}^p\}$ la aplicación polinomial $\tilde{F}^{[p]}$ que a cada $(z, \tilde{z})$ le asigna $(F(z), \tilde{z})$. Así, es fácil ver que F y $\tilde{F}^{[p]}$ son simultáneamente inyectivas (resp. Keller) o no lo son. Del mismo modo se puede definir

las aplicaciones $E_1(z,\tilde{z}) = (z+\tilde{z},\tilde{z})$ y $E_2(z,\tilde{z}) = (z,\tilde{z}-F_{(3)}(z))$. Por esta razón, la inyectividad de F es equivalente a inyectividad de $G = E_1 \circ \tilde{F}^{[p]} \circ E_2$, la misma que satisface

$$\begin{aligned} G(z,\tilde{z}) &= E_1 \circ \tilde{F}^{[p]}(z,\tilde{z}-F_{(3)}(z)), \\ &= E_1\big(F(z),\tilde{z}-F_{(3)}(z)\big), \\ &= (z+F_{(2)}(z)+\tilde{z},\tilde{z}-F_{(3)}(z)), \\ &= (z,\tilde{z})+\big(\tilde{z}+F_{(2)}(z),-F_{(3)}(z)\big). \end{aligned}$$

Además, su matriz jacobiana está dada por

$$DG(z,\tilde{z}) = \begin{pmatrix} I_p + DF_{(2)}(z) & I_p \\ -DF_{(3)}(z) & I_p \end{pmatrix},$$

y así los determinantes de $DG(z,\tilde{z})$ y $DF(z)$ son iguales. Por tanto, *si $F=(F_1,\ldots,F_n)$ es Keller, siempre existe $\hat{m}\geq 0$ tal que la inyectividad de F es equivalente a la inyectividad de alguna aplicación de la forma*

$$z \mapsto z + \tilde{H}_{(1)}(z) + \tilde{H}_{(2)}(z) + \tilde{H}_{(3)}(z),$$

donde $z\in\mathbb{C}^{n+\hat{m}}$, cada componente de $\tilde{H}_{(i)}$ o es cero o es homogéneo de grado $i\in\{1,2,3\}$ y cualquier matriz jacobiana de $\tilde{H}_{(1)}(z)+\tilde{H}_{(2)}(z)+\tilde{H}_{(3)}(z)$ es nilpotente[1].

Observación 4.7. En el acápite § 4.6 no es necesario considerar aplicaciones de Keller: basta elegir una aplicación polinomial cuya matriz jacobiana en el cero tiene inversa y concluir que la matriz jacobiana de $\tilde{H}_{(1)}(z)+\tilde{H}_{(2)}(z)+\tilde{H}_{(3)}(z)$ en el origen es nilpotente.

4.8. La aplicación $\tilde{z}\mapsto \tilde{z}+\tilde{H}_{(1)}(\tilde{z})+\tilde{H}_{(2)}(\tilde{z})+\tilde{H}_{(3)}(\tilde{z})$ induce en el producto $\mathbb{C}^q\times\mathbb{C}=\{(\tilde{z},t):\tilde{z}\in\mathbb{C}^q,t\in\mathbb{C}\}$ la aplicación

$$(\tilde{z},t)\mapsto(\tilde{z},t)+(t^2\tilde{H}_{(1)}(\tilde{z})+t\tilde{H}_{(2)}(\tilde{z})+\tilde{H}_{(3)}(\tilde{z}),0),$$

[1] En el anillo graduado de las matrices cuadradas de orden $n+\hat{m}$ se cumple que $I+a$ tiene inversa si y sólo si $a^N=0$ para algun $N\geq 2$ (i.e., a es nilpotente). De hecho, la inversa será $\sum_{k\geq 0}(-a)^k$. En efecto, cuando la inversa de $I+a$ (con $a\neq 0$) este bien definida, esta será de la forma $I+b_{i_1}+\cdots+b_{i_m}$, y asi la igualdad $(I+a)(I+b_{i_1}+\cdots+b_{i_m})=I$ junto a un natural argumento de inducción permite obtener la igualdad $b_{i_m}=(-a)^m$ y, así, $a^{m+1}=0$.

y ambas son mutuamente inyectivas (resp. Keller). En consecuencia, cuando la aplicación $F = (F_1, \ldots, F_n)$ es Keller, siempre existe $m \geq 0$ tal que la inyectividad de F es equivalente a la inyectividad de alguna aplicación de la forma $\hat{z} \mapsto \hat{z} + H(\hat{z})$, donde $\hat{z} \in \mathbb{C}^{n+m}$, cada componente de H es homogénea de grado tres y la matriz jacobiana de $H(\hat{z})$ es nilpotente. Por tanto, *para dar una respuesta a la conjetura jacobiana de Keller basta estudiar la inyectividad de todas las aplicaciones de la forma* $z \mapsto z + H(z)$*, donde las componentes no nulas de* H *son polinomios homogéneos de grado tres y cada* $DH(z)$ *es nilpotente [2].*

Observación 4.9. El resultado demostrado en § **4.8** aún se puede refinar. Por ejemplo, los autores de [7] muestran que en § **4.8** es posible suponer que cada $DH(z)$ no solo es nilpotente sino también simétrica. Del mismo modo, los trabajos de Drużkowski [8] reducen el estudio de la conjetura jacobiana de Keller a las aplicaciones de la forma

$$\mathbb{C}^n \ni z \mapsto z + \left(\ell_1^3, \ldots, \ell_n^3\right) \in \mathbb{C}^n,$$

donde cada $\ell_j = a_{1,j}z_1 + a_{2,j}z_2 + \cdots + a_{n,j}z_n$ es lineal en $z = (z_1, \ldots, z_n)$.

Apéndice

La siguiente proposición, tomada de [21], fue utilizada en § **2.4**.

Proposición 4.10. *Si* $A = \mathbb{Z}[x_1, \ldots, x_n]$ *es una* $\mathbb{Z}$ *subálgebra de una* $\mathbb{Q}-$*álgebra, entonces existe un conjunto finito* $E \subset \mathbb{Z}$ *fuera del cual cada primo* $p \in \mathbb{Z}$ *está en algún ideal maximal* $\mathfrak{m}_p$ *de* A *y el cociente* $A/\mathfrak{m}_p$ *es un cuerpo finito.*

Prueba. Se procede por indución sobre m. Si $m = 0$ se tiene $A = \mathbb{Z}$ y el resultado se logra con el ideal generado por p. Si $m \geq 1$, se consideran dos casos: cuando $x_1, \ldots, x_n$ son algebraicamente independientes sobre $\mathbb{Q}$, o en su defecto cuando existe un polinomio $f \in \mathbb{Z}[X_1, \ldots, X_n] - \mathbb{Z}$ con $f(x_1, \ldots, x_n) = 0$.

En el primer caso, la proposición se obtiene usando $(p, x_1, \ldots, x_n) \subset A$, el ideal generado por $x_1, \ldots, x_n$ y p.

En el segundo caso, se asume que X_n aparece en la expresión de f y se observa que por cada $N \geq 1$ las reglas $X_n \mapsto X_n$ y $X_j \mapsto X_j + (X_n)^{N^j}$, $j < n$, definen de forma natural un $\mathbb{Z}-$automorfismo ϕ_n de $\mathbb{Z}[X_1, \ldots, X_n]$. Más aún, se obtiene la siguiente propiedad.

Afirmación: *Existe $N \geq 1$ tal que la imagen de f bajo ϕ_n es un polinomio de la forma*

$$cX_n^m + f_{m-1}(X_1, \ldots, X_{n-1})X_n^{m-1} + \cdots + f_0(X_1, \ldots, X_{n-1}),$$

donde $c \in \mathbb{Z} \setminus \{0\}$, y para $0 \leq i \leq m-1$ se tiene $f_i \in \mathbb{Z}[X_1, \ldots, X_{n-1}]$.

Prueba de la afirmación. En efecto, basta elegir $N \geq 1$ como una cota superior de todos los exponentes en cada variable que aparece en f y estudiar los exponentes de X_n, en las imágenes de los monomios que aparecen en la expresión de f. Notemos que estos exponentes se escriben como expresiones polinomiales de N. Por los tanto, la afirmación es verdadera. □

Hecho esto, aplicamos la hipótesis inductiva a $B = \mathbb{Z}[X_1, \ldots, X_{n-1}]$. De este modo, por cada $p \in \mathbb{Z}$ fuera de un conjunto finito $E_1 \subset \mathbb{Z}$, existe un ideal maximal $\eta_p = \eta$ de B de modo que $p \in \eta_p$ y el cociente B/η_p es un cuerpo finito.

Análogamente, para E_2, el conjunto de los divisores primos de $c \in \mathbb{Z}$, definido en la afirmación, se elige $E = E_1 \cup E_2$ y se cumple la propiedad deseada. Para ver esto se usa el proceso de localización, es decir se considera el conjunto $S = \{1, c, c^2, \ldots\} \subset \mathbb{Z} \subset B$ y el anillo $S^{-1}B = \{\frac{g}{\mathfrak{s}} \in \mathbb{Q}[X_1, \ldots . X_{n-1}] : g \in B, \mathfrak{s} \in S\}$. En ese sentido, la afirmación muestra que x_n es la raíz de un polinomio mónico con coeficientes en $S^{-1}B$.

Del mismo modo, el conjunto $\tilde{S} \subset B/\eta$ de las clases de equivalencias inducidas por S (cerrado por la multiplicación) permite definir el anillo $\tilde{S}^{-1}(B/\eta)$. Este anillo es isomorfo al cociente $S^{-1}B/S^{-1}\eta$, donde

$S^{-1}\eta = \{h/s : h \in \eta, s \in S\}$ es un ideal en $S^{-1}B$. Pero $p \in \eta$ y c son primos relativos, y así la clase $\tilde{c} \in B/\eta$ es una unidad en B/η. En consecuencia $\tilde{S}^{-1}(B/\eta)$ es igual a B/η. Por lo tanto $S^{-1}B/S^{-1}\eta \simeq \tilde{S}^{-1}(B/\eta)$ es un cuerpo finito.

Por otro lado se tiene $S^{-1}B \subset S^{-1}B[x_n]$. Como x_n es la raíz de un polinomio mónico con coeficientes en $S^{-1}B$, existe un ideal maximal $\mathfrak{m} \subset S^{-1}B[x_n]$ que contiene a $S^{-1}\eta$. A partir de esto se obtiene que la inclusión $S^{-1}B/S^{-1}\eta \hookrightarrow S^{-1}B[x_n]/\mathfrak{m}$ es integral y por eso finita. Por lo tanto $S^{-1}B[x_n]/\mathfrak{m}$ es un cuerpo finito.

Para terminar consideramos $\mathfrak{P} = \varphi^{-1}(\mathfrak{m})$, donde $\varphi : B[x_n] \to S^{-1}B[x_n]$ es la aplicación canónica. Este $\mathfrak{P}$ es un ideal primo en $B[x_n]$ y así la aplicación $B[x_n]/\mathfrak{P} \mapsto S^{-1}B[x_n]/\mathfrak{m}$ será inyectiva. En consecuencia, $B[x_n]/\mathfrak{P}$ es finito. Por lo tanto, si $\mathfrak{m}_p$ es un ideal maximal de $A = B[x_n]$ que contiene $\mathfrak{P}$, se obtiene que $A/\mathfrak{P} \mapsto A/\mathfrak{m}_p$ es sobreyectiva, y por lo tanto $A/\mathfrak{m}_p$ será un cuerpo finito. □

Referencias

[1] Andrzej Białynicki-Birula and Maxwell Rosenlicht; *Injective morphisms of real algebraic varieties*, Proc. Amer. Math. Soc. **13** (1962), 200–203. MR 0140516 (25 #3936)

[2] Hyman Bass, Edwin H. Connell, and David Wright; *The Jacobian conjecture: reduction of degree and formal expansion of the inverse*, Bull. Amer. Math. Soc. (N.S.) **7** (1982), no. 2, 287–330. MR 663785(83k:14028)

[3] L. Andrew Campbell; *Unipotent Jacobian matrices and univalent maps*, Combinatorial and computational algebra (Hong Kong, 1999), Contemp. Math., vol. **264**, Amer. Math. Soc., Providence, RI, 2000, pp. 157–177. MR 1800694 (2001m:26027)

[4] Eduardo Cabral Balreira; *Foliations and global inversion*, Comment. Math. Helv. **85** (2010), no. 1, 73–93. MR 2563681 (2010j:58085)

[5] Marc Chamberland and Gary Meisters; *A mountain pass to the Jacobian conjecture*, Canad. Math. Bull. **41** (1998), no. 4, 442–451. MR 1658243 (99m:58044)

[6] Sławomir Cynk and Kamil Rusek; *Injective endomorphisms of algebraic and analytic sets*, Ann. Polon. Math. **56** (1991), no. 1, 29–35. MR 1145567 (93d:14025)

[7] Michiel de Bondt and Arno van den Essen; *A reduction of the Jacobian conjecture to the symmetric case*, Proc. Amer. Math. Soc. **133** (2005), no. 8, 2201–2205. MR 2138860 (2006a:14107)

[8] Ludwik M. Drużkowski; *An effective approach to Keller's Jacobian conjecture*, Math. Ann. **264** (1983), no. 3, 303–313. MR 714105 (85b:14015a)

[9] Alexandre Fernandes, Carlos Gutierrez, and Roland Rabanal; *On local diffeomorphisms of $\mathbb{R}^n$ that are injective*, Qual. Theory Dyn. Syst. **4** (2003), no. 2, 255–262 (2004). MR 2129721 (2005m:58021)

[10] A. V. Jagžev; *On a problem of O.-H. Keller*, Sibirsk. Mat. Zh. **21** (1980), no. 5, 141–150, 191. MR 592226 (82e:14020)

[11] Ming Chang Kang; *Injective morphisms of affine varieties*, Proc. Amer. Math. Soc. **119** (1993), no. 1, 1–4. MR 1146862 (93k:14006)

[12] Ott-Heinrich Keller; *Ganze Cremona-Transformationen*, Monatsh. Math. Phys. **47** (1939), no. 1, 299–306. MR 1550818

[13] Serge Lang; *Algebra*, third ed., Graduate Texts in Mathematics, vol. 211, Springer-Verlag, New York, 2002. MR 1878556 (2003e:00003)

[14] Masayoshi Nagata; *Local rings*, Interscience Tracts in Pure and Applied Mathematics, No. 13, Interscience Publishers a division of John Wiley & Sons New York-London, 1962. MR 0155856 (27 #5790)

[15] Scott Nollet and Frederico Xavier; *Global inversion via the Palais-Smale condition*, Discrete Contin. Dyn. Syst. **8** (2002), no. 1, 17–28. MR 1877825 (2002m:58013)

[16] S. Nollet and F. Xavier; *Holomorphic injectivity and the Hopf map*, Geom. Funct. Anal. **14** (2004), no. 6, 1339–1351. MR 2135170 (2006f:32025)

[17] Sergey Pinchuk; *A counterexample to the strong real Jacobian conjecture*, Math. Z. **217** (1994), no. 1, 1–4. MR 1292168 (95g:14018)

[18] Patrick J. Rabier; *Ehresmann fibrations and Palais-Smale conditions for morphisms of Finsler manifolds*, Ann. of Math. (2) **146** (1997), no. 3, 647–691. MR 1491449 (98m:58020)

[19] Roland Rabanal; *On differentiable area-preserving maps of the plane*, Bull. Braz. Math. Soc. (N.S.) **41** (2010), no. 1, 73–82. MR 2609212 (2011e:26011)

[20] Brian Smyth and Frederico Xavier; *Injectivity of local diffeomorphisms from nearly spectral conditions*, J. Differential Equations **130** (1996), no. 2, 406–414. MR 1410896 (98b:58025)

[21] Arno van den Essen; *Polynomial automorphisms and the Jacobian conjecture*, Progress in Mathematics, vol. 190, Birkhäuser Verlag, Basel, 2000. MR 1790619 (2001j:14082)

[22] Arno van den Essen, Ha Huy Vui, Hanspeter Kraft, Peter Russell, and David Wright; *Polynomial automorphisms and related topics*, Publishing House for Science and Technology, Hanoi, 2007, Lecture notes from the International School and Workshop (ICPA2006) held in Hanoi, October 9–20, 2006, Edited by Hyman Bass, Nguyen Van Chau and Stefan Maubach. MR 2389268 (2009a:14001)

[23] Stuart Sui Sheng Wang; *A Jacobian criterion for separability*, J. Algebra **65** (1980), no. 2, 453–494. MR 585736 (83e:14010)

Abstract

The polynomial maps whose Jacobian determinant is equal to 1 are called Keller maps. The Keller Jacobian conjecture claims that every Keller map is injective. This conjecture is true for polynomials whose degree is less than or equal to two. In this paper we prove that the general case reduces to the study of the injectivity of maps of the form $z \mapsto z + H(z)$, where the nonzero components of H are homogeneous polynomials of degree three, and every Jacobian matrix $DH(z)$ is nilpotent.

Keywords: Keller maps, polinomial ring, automorphism.

Percy Fernandez
Sección Matemáticas
Departamento de Ciencias
Pontificia Universidad Católica del Perú
Av. Universitaria 1801, San Miguel, Lima 32, Perú
pefernan@pucp.edu.pe

Roland Rabanal
Sección Matemáticas
Departamento de Ciencias
Pontificia Universidad Católica del Perú
Av. Universitaria 1801, San Miguel, Lima 32, Perú
rrabanal@pucp.edu.pe

Cálculo exacto de la matriz exponencial

Rubén Agapito[1]

Octubre, 2014

Resumen

Presentamos varios métodos que permiten el cálculo exacto de la matriz exponencial e^{tA}. Los métodos que incluyen el cálculo de autovectores y la transformada de Laplace son bien conocidos, y son mencionados aquí por completitud. Se mencionan otros métodos, no tan conocidos en la literatura, que no incluyen el cálculo de autovectores, y que proveen de fórmulas genéricas aplicables a cualquier matriz.

MSC(2010): 15A16, 15A18, 34-01, 44A10.

Palabras Clave: Matriz exponencial, matriz diagonalizable, forma canónica de Jordan, triangularización de Schur, funciones matriciales, interpolación de Lagrange-Sylvester, fórmula espectral de Putzer, transformada de Laplace.

1. *Sección Matemáticas, Departamento de Ciencias, PUCP.*

1. Introducción

La matriz exponencial es una herramienta muy útil en la resolución de sistemas lineales de primer orden. Nos provee de una fórmula cerrada para sus soluciones, y con ayuda de ésta puede analizarse la controlabilidad y observabilidad de un sistema lineal ([1]). Existen varios métodos para calcular la matriz exponencial, ninguno de ellos computacionalmente eficiente ([9]). Sin embargo, desde el punto de vista teórico es importante conocer propiedades de esta función matricial. Fórmulas que involucran el cálculo de autovectores generalizados y transformada de Laplace han sido utilizados en una amplia cantidad de libros texto, y por este motivo, en este trabajo, se pretende brindar métodos alternativos, no muy conocidos, de didáctica amable. Existen otros métodos ([4],[5]) de por sí interesantes pero que no han sido mencionados en la lista de casos, debido a su practicidad en implementación.

Desarrollaremos ocho casos o métodos para calcular la matriz exponencial. Se brindan ejemplos de cómo aplicar los métodos no tan conocidos en casos concretos, y para los casos más conocidos se cita la respectiva bibliografía.

2. Definiciones y resultados básicos

La matriz exponencial e^{tA} puede ser definida generalizando la noción de serie de Maclaurin de la función exponencial escalar vía

$$\Phi(t) = \mathrm{e}^{tA} = \sum_{k=0}^{\infty} \frac{1}{k!} t^k A^k.$$

Para establecer la convergencia de esta serie, definamos primero la norma de Frobenius de una matriz de tamaño $m \times n$ mediante

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}.$$

Si $A(:,j)$ denota a la j-ésima columna de A, y $A(i,:)$ su i-ésima fila, es fácil ver que se cumple

$$\|A\|_F = \left(\sum_{j=1}^{n} \|A(:,j)\|_2^2\right)^{1/2} = \left(\sum_{i=1}^{m} \|A(i,:)\|_2^2\right)^{1/2}.$$

Usaremos esta norma por comodidad, ya que en un espacio vectorial de dimensión finita todas las normas son equivalentes.

Una propiedad importante que utilizaremos es saber cómo acota la norma de Frobenius a un producto de matrices. Dadas las matrices $A_{m\times p}$ y $B_{p\times n}$, formamos el producto $C = AB$, con entradas $c_{ij} = A(i,:)\,B(:,j)$. Si A tuviese entradas complejas, en la obtención de c_{ij} se aplica conjugada a la fila $A(i,:)$. Recuérdese la desigualdad de Cauchy-Schwarz

$$|c_{ij}| \leq \|A(i,:)\|_2 \, \|B(:,j)\|_2.$$

Luego se tiene

$$\begin{aligned}\|AB\|_F^2 &= \sum_{i=1}^{m}\sum_{j=1}^{n} |c_{ij}|^2 \leq \sum_{i=1}^{m}\sum_{j=1}^{n} \|A(i,:)\|_2^2 \|B(:,j)\|_2^2 \\ &= \sum_{i=1}^{m} \|A(i,:)\|_2^2 \sum_{j=1}^{n} \|B(:,j)\|_2^2 = \|A\|_F^2 \|B\|_F^2.\end{aligned}$$

Si aplicamos esta desigualdad a una matriz cuadrada A es fácil deducir lo siguiente

$$\|A^n\|_F \leq \|A\|_F^n, \quad \text{para todo } n = 1, 2, 3, \ldots$$

Formalmente debemos examinar la convergencia del siguiente límite

$$\lim_{n\to\infty} \left(\sum_{k=0}^{n} \frac{A^k}{k!}\right).$$

Para ello basta observar que se satisface

$$\left\|\sum_{k=0}^{n} \frac{A^k}{k!}\right\|_F \leq \sum_{k=0}^{n} \frac{\|A^k\|_F}{k!} \leq \sum_{k=0}^{n} \frac{\|A\|_F^k}{k!} \leq \mathrm{e}^{\|A\|_F},$$

y así queda demostrado que e^A está bien definido para cualquier matriz cuadrada con entradas constantes.

Es útil recordar cómo se comporta la matriz exponencial bajo derivación:

$$\begin{aligned}\Phi^{(1)}(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{k=0}^{\infty}\frac{1}{k!}t^kA^k\right)\\ &= \frac{\mathrm{d}}{\mathrm{d}t}\left(I+tA+\frac{1}{2!}t^2A^2+\cdots+\frac{1}{k!}t^kA^k+\cdots\right)\\ &= A+\frac{2}{2!}tA^2+\cdots+\frac{k}{k!}t^{k-1}A^k+\cdots\\ &= A\left(I+tA+\cdots+\frac{1}{(k-1)!}t^{k-1}A^{k-1}+\cdots\right)=A\,\mathrm{e}^{tA}=\mathrm{e}^{tA}\,A.\end{aligned}$$

Usando inducción y el convenio $\Phi^{(0)}(t)=\Phi(t)$, se deduce la fórmula

$$\Phi^{(k)}(t)=\frac{\mathrm{d}^k}{\mathrm{d}t^k}\,\mathrm{e}^{tA}=A^k\,\mathrm{e}^{tA}=\mathrm{e}^{tA}\,A^k,\quad k\in\mathbb{Z}_0^+. \tag{2.1}$$

Obsérvese que la fórmula para la primera derivada implica que la función $x(t)=\mathrm{e}^{tA}\,x_0$ es solución del problema de valor inicial del siguiente sistema de primer orden

$$\dot{x}=Ax,\qquad x(0)=x_0.$$

Dos resultados conocidos de álgebra lineal (ver [8]) que usaremos más adelante son los siguientes teoremas.

Teorema 2.1 (de triangularización de Schur)**.** *Toda matriz $A_{n\times n}$ es (unitariamente) similar a una matriz triangular superior T, esto es, existe una matriz unitaria U tal que $A=UTU^{-1}$. Además, las entradas en la diagonal de T son los autovalores de A.* □

Teorema 2.2 (de Cayley-Hamilton)**.** *Cualquier matriz $A_{n\times n}$ es raíz de su propio polinomio característico.* □

Pasamos ahora a detallar ocho casos o métodos para hallar la matriz exponencial.

3. Matriz diagonalizable

Dada una matriz diagonal $n \times n$

$$D = \operatorname{diag}(\lambda_1, \lambda_2 \ldots, \lambda_n),$$

es fácil deducir que se cumple $D^k = \operatorname{diag}(\lambda_1^k, \ldots, \lambda_n^k)$ para todo $k \in \mathbb{Z}^+$. Luego

$$\begin{aligned} \mathrm{e}^{tD} &= \sum_{k=0}^{\infty} \frac{t^k}{k!} D^k = \operatorname{diag}\left(\sum_{k=0}^{\infty} \frac{t^k}{k!} \lambda_1^k, \ldots, \sum_{k=0}^{\infty} \frac{t^k}{k!} \lambda_n^k \right) \\ &= \operatorname{diag}\left(\mathrm{e}^{t\lambda_1}, \ldots, \mathrm{e}^{t\lambda_n} \right), \end{aligned}$$

es también una matriz diagonal. Ahora, en el caso que A sea una matriz diagonalizable se sabe que existe una matriz invertible P formada por los autovectores de A y una matriz diagonal D formada por los correspondientes autovalores de A tales que $A = PDP^{-1}$. Es sencillo verificar la identidad $A^k = PD^kP^{-1}$ para todo $k \in \mathbb{Z}^+$. Luego se tiene

$$\begin{aligned} \mathrm{e}^{tA} &= \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k = \sum_{k=0}^{\infty} \frac{t^k}{k!} \left(PD^kP^{-1} \right) = P \left(\sum_{k=0}^{\infty} \frac{t^k}{k!} D^k \right) P^{-1} \\ &= P\, \mathrm{e}^{tD}\, P^{-1}. \end{aligned}$$

En consecuencia, es trivial encontrar la matriz exponencial de una matriz diagonalizable, siempre y cuando hallemos previamente todos los autovalores de A con sus correspondientes autovectores.

Este caso es bien conocido. Ver [11] para apreciar algunos ejemplos de cerca.

4. Matriz no diagonalizable

Cuando una matriz no es diagonalizable se sabe que es similar a una matriz en forma canónica de Jordan, esto es, una matriz diagonal por

bloques, donde cada bloque es de la forma (para un bloque de tamaño $k \times k$ con autovalor λ)

$$J = \begin{bmatrix} \lambda & 1 & & & & \\ & \lambda & 1 & & 0 & \\ & & \lambda & 1 & & \\ & & & \ddots & \ddots & \\ & 0 & & & \lambda & 1 \\ & & & & & \lambda \end{bmatrix}.$$

En este caso la matriz exponencial toma la forma

$$\mathrm{e}^{tJ} = \mathrm{e}^{t\lambda} \begin{bmatrix} 1 & t & t^2/2! & t^3/3! & \cdots & t^{k-1}/(k-1)! \\ & 1 & t & t^2/2! & \cdots & t^{k-2}/(k-2)! \\ & & 1 & t & \cdots & t^{k-3}/(k-3)! \\ & & & \ddots & & \vdots \\ & 0 & & & & t \\ & & & & & 1 \end{bmatrix}.$$

Una manera de demostrar esta fórmula es considerar el sistema de primer orden $\dot{x} = Jx$ con condición inicial $x(0) = x_0$. Por un lado, sabemos que la solución de este sistema está dada por $x(t) = \mathrm{e}^{tJ}\, x_0$. Por otro lado, este sistema es fácil de resolver, empezando por la última ecuación, la cual está desacoplada, y luego se resuelve cada ecuación lineal de primer orden una por una, vía el método del factor integrante.

En el caso que la matriz $A_{n\times n}$ no es diagonalizable, se sabe que es similar a una forma canónica de Jordan $n \times n$

$$J = \mathrm{diag}(J_1, J_2, \ldots, J_m),$$

en donde hemos supuesto que existen m autovalores distintos con multiplicidades mayores o iguales a uno. Si algún bloque de Jordan J_k, $k = 1, 2, \ldots, m$, es de tamaño 2×2 o superior, se debe a que el autovalor λ_k no posee una base completa de autovectores, y debe entonces

completarse con autovectores generalizados, ver [11]. Si formamos la matriz P en donde cada columna es un autovector (generalizado), resulta ser una matriz invertible y se cumple

$$A = PJP^{-1}.$$

Al aplicar un cálculo similar al caso diagonalizable se demuestra

$$\mathrm{e}^{tA} = P\,\mathrm{diag}\big(\mathrm{e}^{tJ_1}, \mathrm{e}^{tJ_2}, \ldots, \mathrm{e}^{tJ_m}\big)P^{-1}.$$

Este caso es también muy conocido. Ver [11] para apreciar algunos ejemplos.

5. Matrices triangulares

Sea S una matriz triangular superior (para una matriz triangular inferior se realiza un desarrollo similar) y escribámosla como la suma de una matriz diagonal con una matriz nilpotente

$$S = D + N.$$

Recuérdese que una matriz N es llamada *nilpotente* si existe un entero positivo r tal que $N^r = 0$. El menor entero positivo para el cual esta igualdad se cumple, es llamado el *índice de nilpotencia* de la matriz.

Asumiendo la conocida propiedad sobre matrices exponenciales (ver [1])

$$\mathrm{e}^{A+B} = \mathrm{e}^{A}\ \mathrm{e}^{B}, \quad \text{si } AB = BA,$$

calculamos

$$\mathrm{e}^{tS} = \mathrm{e}^{t(D+N)} = \mathrm{e}^{tD}\,\mathrm{e}^{tN}\,.$$

Sabemos cómo calcular la matriz exponencial de una matriz diagonal, así que discutamos cómo obtener e^{tN}. Basta observar que al ser N nilpotente, la serie de esta matriz se vuelve finita, ya que el número de sumandos queda acotado por el índice de nilpotencia de N. Éste a su

vez queda acotado por el grado de su polinomio minimal (recordar que toda matriz nilpotente posee todos sus autovalores iguales a cero).

Podemos generalizar este método a cualquier matriz $A_{n\times n}$. Para ello, basta aplicar el teorema de triangularización de Schur, $A = USU^{-1}$, donde U es una matriz unitaria y S está en forma triangular superior. Con ello obtenemos

$$\mathrm{e}^{tA} = U\,\mathrm{e}^{tS}\,U^{-1},$$

y de aquí ya sabemos cómo proceder.

6. La fórmula espectral de Putzer

En [10], Putzer describe dos métodos para calcular e^{tA}. Estos se basan en el hecho de que e^{tA} es un polinomio en A cuyos coeficientes son funciones escalares de t que pueden ser halladas recursivamente resolviendo un sistema sencillo de ecuaciones diferenciales lineales de primer orden. Mostraremos sólo el segundo método, por ser más fácil de entender e implementar.

Teorema 6.1. *Dada una matriz A de tamaño $n\times n$, supongamos que conocemos todos sus autovalores $\lambda_1, \lambda_2, \ldots, \lambda_n$, no necesariamente distintos, listados en un orden especificado pero arbitrario. Entonces se cumple*

$$\mathrm{e}^{tA} = r_1(t)P_0 + r_2(t)P_1 + \cdots + r_n(t)P_{n-1},$$

donde

$$P_0 = I, \qquad P_k = \prod_{j=1}^{k}(A - \lambda_j I), \quad k = 1, 2, \ldots, n-1,$$

y $r_1(t), \ldots, r_n(t)$ son soluciones del sistema diferencial

$$\begin{aligned}
\dot{r}_1 &= \lambda_1 r_1, & r_1(0) &= 1,\\
\dot{r}_2 &= \lambda_2 r_2 + r_1, & r_2(0) &= 0,\\
&\vdots & &\\
\dot{r}_n &= \lambda_n r_n + r_{n-1}, & r_n(0) &= 0.
\end{aligned}$$

Prueba. Ponemos

$$\Phi(t) = \sum_{k=0}^{n-1} r_{k+1}(t) P_k,$$

y definimos $r_0(t) \equiv 0$. Teniendo en cuenta $\dot{r}_{k+1} = \lambda_{k+1} r_{k+1} + r_k$, calculamos

$$\begin{aligned}\dot{\Phi}(t) - \lambda_n \Phi(t) &= \sum_{k=0}^{n-1} \dot{r}_{k+1}(t) P_k - \sum_{k=0}^{n-1} \lambda_n r_{k+1}(t) P_k \\ &= \left(\sum_{k=0}^{n-1} \lambda_{k+1} r_{k+1} P_k + \sum_{k=0}^{n-1} r_k P_k \right) - \left(\sum_{k=0}^{n-2} \lambda_n r_{k+1} P_k + \lambda_n r_n P_{n-1} \right).\end{aligned}$$

Como del primer término podemos extraer $\lambda_n r_n P_{n-1}$, y el segundo puede ser reescrito como

$$\sum_{k=0}^{n-1} r_k P_k = \sum_{k=1}^{n-1} r_k P_k = \sum_{k=0}^{n-2} r_{k+1} P_{k+1},$$

el lado derecho de la igualdad se simplifica a

$$\sum_{k=0}^{n-2} \Big[(\lambda_{k+1} - \lambda_n) P_k + P_{k+1} \Big] r_{k+1}.$$

Ahora, como se cumple $P_{k+1} = (A - \lambda_{k+1} I) P_k$, la expresión entre corchetes se reduce a $(A - \lambda_n I) P_k$. Además se tiene

$$\begin{aligned}\sum_{k=0}^{n-2} (A - \lambda_n I) P_k r_{k+1} &= \sum_{k=0}^{n-1} (A - \lambda_n I) P_k r_{k+1} - \underbrace{(A - \lambda_n I) P_{n-1}}_{P_n} r_n, \\ &= (A - \lambda_n I)\Phi - P_n r_n.\end{aligned}$$

Pero el teorema de Cayley-Hamilton fuerza $P_n = 0$. Así, hemos obtenido que $\dot{\Phi} - \lambda_n \Phi = (A - \lambda_n I)\Phi$ implica $\dot{\Phi} = A\Phi$. Por último, como $\Phi(0) = r_1(0) P_0 = I$, se sigue $\Phi(t) = \mathrm{e}^{tA}$ por unicidad de soluciones. □

Como aplicación de este método, hallemos fórmulas para la matriz exponencial de una matriz 2×2 en la forma

$$\mathrm{e}^{tA} = r_1(t)I + r_2(t)(A - \lambda_1 I),$$

con autovalores $\{\lambda_1, \lambda_2\}$. De acuerdo con la naturaleza de los autovalores tenemos tres casos a estudiar.

- **Autovalores reales y distintos.** Debemos resolver el sistema de ecuaciones

$$\begin{aligned} \dot{r}_1 &= \lambda_1 r_1, & r_1(0) &= 1, \\ \dot{r}_2 &= \lambda_2 r_2 + r_1, & r_2(0) &= 0. \end{aligned}$$

  Si resolvemos la primera ecuación, la cual siempre está desacoplada, obtenemos $r_1(t) = \mathrm{e}^{\lambda_1 t}$. Para la segunda, usando el método del factor integrante obtenemos

$$r_2(t) = \frac{1}{\lambda_1 - \lambda_2}\,\mathrm{e}^{(\lambda_1+\lambda_2)t} - \frac{1}{\lambda_1 - \lambda_2}\,\mathrm{e}^{\lambda_2 t}\,.$$

  En consecuencia, logramos la siguiente fórmula

$$\mathrm{e}^{tA} = \mathrm{e}^{\lambda_1 t}\, I + \frac{\mathrm{e}^{\lambda_2 t}}{\lambda_1 - \lambda_2}\big(\mathrm{e}^{\lambda_1 t} - 1\big)(A - \lambda_1 I).$$

- **Autovalores reales e iguales.** En este caso, al resolver el sistema de ecuaciones obtenemos $r_1(t) = \mathrm{e}^{\lambda_1 t}$ y $r_2(t) = t\,\mathrm{e}^{\lambda_1 t}$. Así, obtenemos la fórmula

$$\mathrm{e}^{tA} = \mathrm{e}^{\lambda_1 t}\, I + t\,\mathrm{e}^{\lambda_1 t}(A - \lambda_1 I).$$

- **Autovalores complejos.** En el caso $A \in \mathbb{C}^{2\times 2}$, con autovalores λ_1, λ_2, no habría problema en utilizar la misma fórmula que en el caso de autovalores reales y distintos. Pero si A tiene entradas reales, sus autovalores serían complejos y conjugados, digamos

$\lambda_1 = a + ib$, $\lambda_2 = a - ib$, con $b \neq 0$. En este caso, al resolver el sistema de ecuaciones obtenemos

$$r_1(t) = \mathrm{e}^{\lambda_1 t} = \mathrm{e}^{at}\big[\cos(bt) + i\,\mathrm{sen}(bt)\big],$$
$$r_2(t) = \frac{\mathrm{e}^{\lambda_1 t} - \mathrm{e}^{\lambda_2 t}}{\lambda_1 - \lambda_2} = \mathrm{e}^{at}\,\frac{\mathrm{sen}(bt)}{b}.$$

Como $x(t) = \mathrm{e}^{tA}\,x_0$ es solución del sistema $\dot{x} = Ax$, con condición inicial $x(0) = x_0$, al tener x_0 y A entradas reales, buscamos obviamente una solución real. Así, la parte real de la fórmula espectral es la solución real a considerar, esto es,

$$x(t) = \mathfrak{Re}\{x(t)\} = \big(\mathfrak{Re}\{r_1(t)\}I + \mathfrak{Re}\{r_2(t)(A - \lambda_1 I)\}\big)x_0 = \mathrm{e}^{tA}\,x_0,$$

y en consecuencia se concluye

$$\mathrm{e}^{tA} = \mathrm{e}^{at}\cos(bt)I + \mathrm{e}^{at}\,\frac{\mathrm{sen}(bt)}{b}(A - aI).$$

7. Los casos particulares de Apostol

En [2], Apostol muestra cómo obtener fórmulas explícitas para la matriz exponencial e^{tA} en los siguientes casos:

- todos los autovalores de A son iguales,
- todos los autovalores de A son distintos,
- A tiene solo dos autovalores distintos, con uno de ellos de multiplicidad algebraica uno.

Si bien estos casos no cubren todas las alternativas posibles para el conjunto de autovalores de una matriz, exhibiremos estas fórmulas por su sencillez y porque nos ayudan a encontrar todas las fórmulas posibles para las matrices exponenciales de tamaño menor o igual a 3×3. Cabe señalar que la fórmula espectral de Putzer también nos ayudaría a deducir estas fórmulas, pero la manera obtenida por Apostol es más contundente.

Teorema 7.1. *Si A es una matriz $n\times n$ con todos sus autovalores iguales a λ entonces tenemos*

$$\mathrm{e}^{tA} = \mathrm{e}^{\lambda t}\sum_{k=0}^{n-1}\frac{t^k}{k!}(A-\lambda I)^k.$$

Prueba. Como las matrices λtI y $t(A-\lambda I)$ conmutan, tenemos

$$\mathrm{e}^{tA} = \mathrm{e}^{\lambda tI}\,\mathrm{e}^{t(A-\lambda I)} = (\mathrm{e}^{\lambda t}\,I)\sum_{k=0}^{\infty}\frac{t^k}{k!}(A-\lambda I)^k.$$

El teorema de Cayley-Hamilton implica $(A-\lambda I)^k = 0$ para $k \geq n$, y así el teorema queda demostrado. □

Teorema 7.2. *Si A es una matriz $n \times n$ con n autovalores distintos $\lambda_1, \lambda_2, \ldots, \lambda_n$, entonces tenemos*

$$\mathrm{e}^{tA} = \sum_{k=1}^{n}\mathrm{e}^{t\lambda_k}\,L_k(A),$$

donde los $L_k(A)$ son los coeficientes de interpolación de Lagrange dados por

$$L_k(A) = \prod_{\substack{j=1\\ j\neq k}}^{n}\frac{A-\lambda_j I}{\lambda_k-\lambda_j} \quad \textit{para } k = 1, 2, \ldots, n.$$

Prueba. Aunque este teorema es un caso especial de la fórmula de interpolación de Lagrange-Sylvester (ver Sección 8, más adelante) daremos una prueba directa.

Definamos la siguiente función matricial de variable escalar

$$F(t) = \sum_{k=1}^{n}\mathrm{e}^{t\lambda_k}\,L_k(A).$$

Para probar $F(t) = \mathrm{e}^{tA}$, mostraremos que F satisface la ecuación diferencial $F'(t) = AF(t)$, con condición inicial $F(0) = I$. En efecto, observemos

que se cumple

$$AF(t) - F'(t) = \sum_{k=1}^{n} \mathrm{e}^{t\lambda_k}(A - \lambda_k I)L_k(A).$$

Por el teorema de Cayley-Hamilton se tiene $(A - \lambda_k I)L_k(A) = 0$ para cada k, y así F satisface la ecuación diferencial. Además, de

$$F(0) = \sum_{k=1}^{n} L_k(A) = I,$$

se deduce finalmente $F(t) = \mathrm{e}^{tA}$ por unicidad de soluciones. □

Teorema 7.3. *Sea A una matriz $n \times n$ ($n \geq 3$) con dos autovalores distintos λ y μ, donde λ tiene multiplicidad $n-1$ y μ tiene multiplicidad 1. Entonces se cumple*

$$\mathrm{e}^{tA} = \mathrm{e}^{\lambda t} \sum_{k=0}^{n-2} \frac{t^k}{k!}(A - \lambda I)^k + \left\{ \frac{\mathrm{e}^{\mu t}}{(\mu - \lambda)^{n-1}} - \frac{\mathrm{e}^{\lambda t}}{(\mu - \lambda)^{n-1}} \sum_{k=0}^{n-2} \frac{t^k}{k!}(\mu - \lambda)^k \right\} (A - \lambda I)^{n-1}.$$

Prueba. En versión escalar, para t fijo, la expansión de e^x en serie de Taylor centrada en λt es

$$\mathrm{e}^x = \sum_{k=0}^{\infty} \frac{\mathrm{e}^{\lambda t}}{k!}(x - \lambda t)^k.$$

Ahora evaluamos en tA y particionamos convenientemente esta serie

$$\begin{aligned} \mathrm{e}^{tA} &= \sum_{k=0}^{\infty} \frac{\mathrm{e}^{\lambda t}}{k!}(tA - \lambda t I)^k = \mathrm{e}^{\lambda t} \sum_{k=0}^{\infty} \frac{t^k}{k!}(A - \lambda I)^k \\ &= \mathrm{e}^{\lambda t} \sum_{k=0}^{n-2} \frac{t^k}{k!}(A - \lambda I)^k + \mathrm{e}^{\lambda t} \sum_{k=n-1}^{\infty} \frac{t^k}{k!}(A - \lambda I)^k \\ &= \mathrm{e}^{\lambda t} \sum_{k=0}^{n-2} \frac{t^k}{k!}(A - \lambda I)^k + \mathrm{e}^{\lambda t} \sum_{i=0}^{\infty} \frac{t^{n-1+i}}{(n-1+i)!}(A - \lambda I)^{n-1+i}. \end{aligned}$$

Ahora reescribimos el segundo sumando de la última expresión. Como se cumple $A - \mu I = A - \lambda I - (\mu - \lambda)I$, tenemos, utilizando el teorema de Cayley-Hamilton, la igualdad

$$0 = (A - \lambda I)^{n-1}(A - \mu I) = (A - \lambda I)^n - (\mu - \lambda)(A - \lambda I)^{n-1},$$

esto es, $(A - \lambda I)^n = (\mu - \lambda)(A - \lambda I)^{n-1}$. Por inducción, se deduce

$$(A - \lambda I)^{n+i} = (\mu - \lambda)^{i+1}(A - \lambda I)^{n-1},$$

y así también

$$\begin{aligned}(A - \lambda I)^{n-1+i} &= (A - \lambda I)^{n-1}(A - \lambda I)^i = \frac{1}{\mu - \lambda}(A - \lambda I)^n(A - \lambda I)^i \\ &= \frac{1}{\mu - \lambda}(A - \lambda I)^{n+i} = (\mu - \lambda)^i(A - \lambda I)^{n-1}.\end{aligned}$$

Al reemplazar esta relación en la segunda suma obtenemos

$$\begin{aligned}&\left\{\sum_{i=0}^{\infty} \frac{t^{n-1+i}}{(n-1+i)!}(\mu - \lambda)^i\right\}(A - \lambda I)^{n-1} = \\ &\qquad = \frac{1}{(\mu - \lambda)^{n-1}}\left\{\sum_{k=n-1}^{\infty} \frac{t^k}{k!}(\mu - \lambda)^k\right\}(A - \lambda I)^{n-1} \\ &\qquad = \frac{1}{(\mu - \lambda)^{n-1}}\left\{\mathrm{e}^{t(\mu-\lambda)} - \sum_{k=0}^{n-2} \frac{t^k}{k!}(\mu - \lambda)^k\right\}(A - \lambda I)^{n-1},\end{aligned}$$

con lo cual culmina la demostración. □

Como aplicación deducimos las fórmulas de la matriz exponencial de cualquier matriz A de tamaño 3×3 de acuerdo con la multiplicidad de sus autovalores.

- Para un autovalor λ con multiplicidad algebraica tres, tenemos

$$\mathrm{e}^{tA} = \mathrm{e}^{\lambda t}\Big\{I + t(A - \lambda I) + \frac{1}{2}t^2(A - \lambda I)^2\Big\}.$$

- Para autovalores distintos $\lambda_1, \lambda_2, \lambda_3$, tenemos

$$\mathrm{e}^{tA} = \mathrm{e}^{\lambda_1 t}\,\frac{(A-\lambda_2 I)(A-\lambda_3 I)}{(\lambda_1-\lambda_2)(\lambda_1-\lambda_3)} + \mathrm{e}^{\lambda_2 t}\,\frac{(A-\lambda_1 I)(A-\lambda_3 I)}{(\lambda_2-\lambda_1)(\lambda_2-\lambda_3)} + \mathrm{e}^{\lambda_3 t}\,\frac{(A-\lambda_1 I)(A-\lambda_2 I)}{(\lambda_3-\lambda_1)(\lambda_3-\lambda_2)}.$$

- Para autovalores distintos λ_1, λ_2 con λ_1 de multiplicidad algebraica dos, tenemos

$$\mathrm{e}^{tA} = \mathrm{e}^{\lambda_1 t}\big\{I + t(A-\lambda_1 I)\big\} + \frac{\mathrm{e}^{\lambda_2 t} - \mathrm{e}^{\lambda_1 t}}{(\lambda_2-\lambda_1)^2}(A-\lambda_1 I)^2 - \frac{t\,\mathrm{e}^{\lambda_1 t}}{\lambda_2-\lambda_1}(A-\lambda_1 I)^2.$$

8. Interpolación de Lagrange-Sylvester y el algoritmo de Gantmacher

El siguiente método, ilustrado en Gantmacher [3], no solo nos ayuda a calcular la matriz exponencial, sino también la evaluación matricial de cualquier función analítica. Primero mencionamos el caso en que los autovalores de una matriz son distintos, y luego el caso general cuando existen multiplicidades.

Teorema 8.1. *Si $f(A)$ es una matriz polinomial en $A_{n\times n}$ y si los autovalores de A son distintos, entonces $f(A)$ puede ser descompuesta como*

$$f(A) = \sum_{i=1}^{n} f(\lambda_i) z(\lambda_i),$$

acá $\{\lambda_i\}_{i=1,\ldots,n}$ son los autovalores de A y $z(\lambda_i)$ es la matriz $n \times n$ dada por

$$z(\lambda_i) = \prod_{\substack{k=1 \\ k\neq i}}^{n} \frac{A-\lambda_k I}{\lambda_i - \lambda_k}.$$

Prueba. Por el teorema de Cayley-Hamilton $f(A)$ puede ser reducida a un polinomio de grado $n-1$, digamos

$$f(A) = c_{n-1}A^{n-1} + c_{n-2}A^{n-2} + \ldots + c_1 A + c_0 I.$$

Esta expresión polinomial puede ser factorizada vía interpolación de Lagrange cual

$$f(A) = \sum_{i=1}^{n} p_i \prod_{\substack{k=1 \\ k\neq i}}^{n} (A - \lambda_k I).$$

Para calcular los pesos p_i, multiplicamos por la derecha a ambos lados de la igualdad por el j-ésimo autovector v_j correspondiente a λ_j. Este procedimiento nos entrega

$$\begin{aligned} f(A)v_j &= \sum_{i=1}^{n} p_i \prod_{\substack{k=1 \\ k\neq i}}^{n} (A - \lambda_k I)v_j = p_j \prod_{\substack{k=1 \\ k\neq j}}^{n} (Av_j - \lambda_k v_j) \\ &= p_j \prod_{\substack{k=1 \\ k\neq j}}^{n} (\lambda_j - \lambda_k)v_j, \end{aligned}$$

donde la segunda igualdad se obtiene al considerar $(A - \lambda_j I)v_j = 0$. Como los autovalores de A son distintos, se tiene $f(A)v_j = f(\lambda_j)v_j$, y se deduce fácilmente por comparación la igualdad

$$p_j = \frac{f(\lambda_j)}{\prod\limits_{\substack{k=1 \\ k\neq j}}^{n} (\lambda_j - \lambda_k)}.$$

Por lo tanto, todo junto nos da

$$f(A) = \sum_{i=1}^{n} f(\lambda_i) \prod_{\substack{k=1 \\ k\neq i}}^{n} \frac{A - \lambda_k I}{\lambda_i - \lambda_k}.$$

□

Cuando los autovalores no son distintos, el algoritmo de Gantmacher, ver [3], puede ser usado para expandir la función analítica $f(A)$ como

$$f(A) = \sum_{j=1}^{m} \sum_{k=1}^{m_j} \frac{f^{(k-1)}(\lambda_j)}{(k-1)!} Z_{jk},$$

donde m es el número de autovalores distintos, m_j es la multiplicidad del j-ésimo autovalor, $f^{(k)}(\lambda_j)$ es la derivada con respecto a λ evaluada en el j-ésimo autovalor y Z_{jk} son las *matrices constituyentes* (constantes) que una vez halladas son fijas para cualquier función analítica $f(A)$.

Ilustremos el uso de esta fórmula con un ejemplo.

Ejemplo 8.2. Hallemos e^{tA} para la matriz

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 1 & 3 & 2 \\ -1 & -1 & 6 \end{bmatrix},$$

con polinomio característico $p_A(\lambda) = (\lambda - 2)(\lambda - 4)^2$. Tenemos entonces dos autovalores distintos: $\lambda_1 = 2$ con multiplicidad $m_1 = 1$ y $\lambda_2 = 4$ con multiplicidad $m_2 = 2$. Luego se tiene

$$\begin{aligned} f(A) &= \sum_{j=1}^{2} \sum_{k=1}^{m_j} \frac{f^{(k-1)}(\lambda_j)}{(k-1)!} Z_{jk} \\ &= \sum_{k=1}^{1} \frac{f^{(k-1)}(\lambda_1)}{(k-1)!} Z_{1k} + \sum_{k=1}^{2} \frac{f^{(k-1)}(\lambda_2)}{(k-1)!} Z_{2k} \\ &= f(\lambda_1) Z_{11} + f(\lambda_2) Z_{21} + f'(\lambda_2) Z_{22}. \end{aligned}$$

Recordemos que esta expansión es válida para cualquier función analítica. En nuestra situación, estamos interesados en la versión matricial de $f(\lambda) = \mathrm{e}^{t\lambda}$. Aquí los coeficientes a considerar son

$$f(\lambda_1) = \mathrm{e}^{2t}, \quad f(\lambda_2) = \mathrm{e}^{4t}, \quad f'(\lambda_2) = t\,\mathrm{e}^{4t}.$$

Para hallar las matrices constituyentes utilizamos funciones polinomiales conocidas. Como la matriz es de tamaño 3×3, el teorema de Cayley-Hamilton afirma que e^{tA} debe ser una función polinomial en I, A, y A^2. Usaremos entonces los siguientes criterios.

- Si $f(A) = I$, la versión escalar es $f(\lambda) = 1$. Luego, los coeficientes son $f(\lambda_1) = 1, f(\lambda_2) = 1, f'(\lambda_2) = 0$. Obtenemos así la ecuación $I = Z_{11} + Z_{21}$.
- Si $f(A) = A$, la versión escalar resulta $f(\lambda) = \lambda$. Los coeficientes son $f(\lambda_1) = 2, f(\lambda_2) = 4, f'(\lambda_2) = 1$. Obtenemos la ecuación $A = 2Z_{11} + 4Z_{21} + Z_{22}$.
- Si $f(A) = A^2$, con versión escalar $f(\lambda) = \lambda^2$, los coeficientes son $f(\lambda_1) = 4, f(\lambda_2) = 16, f'(\lambda_2) = 8$. Obtenemos entonces la ecuación $A^2 = 4Z_{11} + 16Z_{21} + 8Z_{22}$.

En consecuencia, debemos resolver el sistema matricial formal

$$\begin{bmatrix} 1 & 1 & 0 \\ 2 & 4 & 1 \\ 4 & 16 & 8 \end{bmatrix} \begin{bmatrix} Z_{11} \\ Z_{21} \\ Z_{22} \end{bmatrix} = \begin{bmatrix} I \\ A \\ A^2 \end{bmatrix},$$

cuya solución es

$$\begin{bmatrix} Z_{11} \\ Z_{21} \\ Z_{22} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 16 & -8 & 1 \\ -12 & 8 & -1 \\ 16 & -12 & 2 \end{bmatrix} \begin{bmatrix} I \\ A \\ A^2 \end{bmatrix} = \begin{bmatrix} 4I - 2A + \frac{1}{4}A^2 \\ -3I + 2A - \frac{1}{4}A^2 \\ 4I - 3A + \frac{1}{2}A^2 \end{bmatrix}.$$

Al poner todo junto vemos que

$$\begin{aligned} f(A) = \mathrm{e}^{tA} &= \mathrm{e}^{2t} Z_{11} + \mathrm{e}^{4t} Z_{21} + t\,\mathrm{e}^{4t} Z_{22} \\ &= \mathrm{e}^{2t} \begin{bmatrix} 9/4 & 9/4 & -3 \\ -7/4 & 1/4 & -3 \\ 9/4 & 9/4 & 1 \end{bmatrix} + \mathrm{e}^{4t} \begin{bmatrix} -5/4 & -9/4 & 3 \\ 7/4 & 3/4 & 3 \\ -9/4 & -9/4 & 0 \end{bmatrix} \\ &\quad + t\,\mathrm{e}^{4t} \begin{bmatrix} 3/2 & 7/2 & -4 \\ -5/2 & -1/2 & -4 \\ 7/2 & 7/2 & 4 \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} \frac{1}{4}e^{2t}\left(e^{2t}(6t-5)+9\right) & \frac{1}{4}e^{2t}\left(e^{2t}(14t-9)+9\right) & e^{4t}(3-4t)-3e^{2t} \\ -\frac{1}{4}e^{2t}\left(e^{2t}(10t-7)+7\right) & \frac{1}{4}\left(e^{4t}(3-2t)+e^{2t}\right) & e^{4t}(3-4t)-3e^{2t} \\ \frac{1}{4}e^{2t}\left(e^{2t}(14t-9)+9\right) & \frac{1}{4}e^{2t}\left(e^{2t}(14t-9)+9\right) & 4e^{4t}t+e^{2t} \end{bmatrix}$$

es la matriz exponencial buscada. Cabe insistir en que una vez halladas las matrices constituyentes, podemos hallar fácilmente $\cos(A), \operatorname{sen}(A), \ldots$, esto es, funciones matriciales de funciones analíticas.

9. Uso de soluciones fundamentales de una ecuación diferencial lineal con coeficientes constantes

El siguiente método evita preguntarse si la matriz es diagonalizable, y tiene como prerrequisitos apenas el conocer el teorema de Cayley-Hamilton y saber cómo resolver ecuaciones diferenciales lineales escalares homogéneas de orden n con coeficientes constantes, digamos tipo

$$x^{(n)} + c_{n-1}x^{(n-1)} + \cdots + c_1x' + c_0x = 0.$$

El método está basado en el artículo de Leonard [6]. Cabe notar que el método no es fácil de aplicar si las raíces de la ecuación polinomial (o ecuación característica) asociada a la ecuación anterior son tediosas de obtener algebraicamente.

Los siguientes teoremas nos ayudarán a entender cómo funciona el método. El primer teorema garantiza la existencia y unicidad de un problema de valor inicial para una ecuación diferencial matricial, mientras el segundo brinda un método para construir la matriz exponencial en base a soluciones de problemas de valor inicial de ecuaciones diferenciales escalares.

Teorema 9.1. *Sea A una matriz constante $n \times n$ con polinomio característico*

$$p(\lambda) = \det(\lambda I - A) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0.$$

Entonces $\Phi(t) = \mathrm{e}^{tA}$ *es la única solución de la ecuación diferencial matricial de orden* n *dada por*

$$\Phi^{(n)} + c_{n-1}\Phi^{(n-1)} + \cdots + c_1\Phi' + c_0\Phi = 0, \tag{9.1}$$

con condición inicial

$$\Phi(0) = I, \quad \Phi'(0) = A, \quad \Phi''(0) = A^2, \ldots, \Phi^{(n-1)}(0) = A^{n-1}. \tag{9.2}$$

Prueba. Primero demostraremos la unicidad. Supongamos que $\Phi_1(t)$ y $\Phi_2(t)$ son dos soluciones de (9.1) que satisfacen las condiciones iniciales dadas en (9.2). Definamos $\Phi(t) = \Phi_1(t) - \Phi_2(t)$. Debido a linealidad, esta función satisface (9.1), pero con condición inicial

$$\Phi(0) = \Phi^{(1)}(0) = \cdots = \Phi^{(n-1)}(0) = 0.$$

Esto implica que cada entrada de $\Phi(t)$ satisface el siguiente problema de valores iniciales

$$\begin{gathered} x^{(n)}(t) + c_{n-1}x^{(n-1)}(t) + \cdots + c_1x^{(1)}(t) + c_0x(t) = 0, \\ x(0) = x^{(1)}(0) = \cdots = x^{(n-1)}(0) = 0, \end{gathered}$$

donde es obvio que la única solución es $x(t) = 0$ para todo t, la trivial. Por ende se tiene $\Phi(t) = 0$ para todo t, y obtenemos unicidad.

Ahora demostramos la existencia al confirmar que $\Phi(t) = \mathrm{e}^{tA}$ satisface el problema de valores iniciales (9.1)-(9.2). Sea A la matriz constante con polinomio característico $p(\lambda)$ descrito en la hipótesis. Recordemos la fórmula de la derivada k-ésima de la exponencial (ver Ecuación (2.1))

$$\Phi^{(k)}(t) = A^k\,\mathrm{e}^{tA}, \quad k = 0, 1, 2, \ldots$$

Reemplazamos en el lado derecho de la Ecuación (9.1) y obtenemos

$$\left(A^n + c_{n-1}A^{n-1} + \cdots + c_1A + c_0I\right)\mathrm{e}^{tA} = p(A)\,\mathrm{e}^{tA} = 0$$

respaldados por el Teorema de Cayley-Hamilton. Por último, de la fórmula de la derivada k-ésima se deduce

$$\Phi^{(0)}(0) = I,\ \Phi^{(1)}(0) = A,\ \ldots, \Phi^{(n-1)}(0) = A^{n-1};$$

así $\Phi(t)$ satisface las condiciones iniciales. Ello termina la demostración. □

Teorema 9.2. *Sea A una matriz constante $n \times n$ con polinomio característico*

$$p(\lambda) = \det(\lambda I - A) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0.$$

Entonces se tiene

$$\mathrm{e}^{tA} = x_1(t)I + x_2(t)A + x_3(t)A^2 + \cdots + x_n(t)A^{n-1},$$

donde los $x_k(t)$, $1 \leq k \leq n$, son las soluciones de las ecuaciones diferenciales escalares de orden n dada por

$$x^{(n)} + c_{n-1}x^{(n-1)} + \cdots + c_1x' + c_0x = 0 \tag{9.3}$$

y que satisfacen las condiciones iniciales

$$\begin{array}{ccc} x_1(0) = 1 & x_2(0) = 0 & x_n(0) = 0 \\ x_1'(0) = 0 & x_2'(0) = 1 & x_n'(0) = 0 \\ \vdots & \vdots & \vdots \\ x_1^{(n-1)}(0) = 0, & x_2^{(n-1)}(0) = 0, & x_n^{(n-1)}(0) = 1. \end{array} \tag{9.4}$$

Prueba. Definamos $\Phi(t) = x_1(t)I + x_2(t)A + x_3(t)A^2 + \cdots + x_n(t)A^{n-1}$, donde los $x_k(t)$ son soluciones de los problemas de valor inicial mencionados en el enunciado. Primero mostraremos que $\Phi(t)$ satisface la Ecuación (9.1). En efecto, se tiene

$$\begin{aligned} \Phi^{(n)}(t) + c_{n-1}\Phi^{(n-1)}(t) + \cdots + c_1\Phi^{(1)}(t) + c_0\Phi(t) \\ &= \big(x_1^{(n)} + c_{n-1}x_1^{(n-1)} + \cdots + c_1x_1' + c_0x_1\big)I \\ &\quad + \big(x_2^{(n)} + c_{n-1}x_2^{(n-1)} + \cdots + c_1x_2' + c_0x_2\big)A + \cdots \\ &\quad + \big(x_n^{(n)} + c_{n-1}x_n^{(n-1)} + \cdots + c_1x_n' + c_0x_n\big)A^{n-1} \\ &= 0I + 0A + \cdots + 0A^{n-1} = 0. \end{aligned}$$

Ahora mostraremos que satisface la condición inicial dada en (9.2):

$$\begin{aligned}
\Phi(0) &= x_1(0)I + x_2(0)A + \cdots + x_n(0)A^{n-1} = I,\\
\Phi'(0) &= x_1'(0)I + x_2'(0)A + \cdots + x_n'(0)A^{n-1} = A,\\
&\vdots\\
\Phi^{(n-1)}(0) &= x_1^{(n-1)}(0)I + x_2^{(n-1)}(0)A + \cdots + x_n^{(n-1)}(0)A^{n-1} = A^{n-1}.
\end{aligned}$$

Luego, por unicidad, se cumple

$$\mathrm{e}^{tA} = \Phi(t) = x_1(t)I + x_2(t)A + \cdots + x_n(t)A^{n-1}$$

para todo t. □

Pasemos a ilustrar el método.

Ejemplo 9.3. Deseamos hallar la matriz exponencial e^{tA} de

$$A = \begin{bmatrix} 1 & -1 & 2\\ 1 & 3 & 2\\ -1 & -1 & 6 \end{bmatrix}.$$

Para ello calculamos su polinomio característico cual

$$p(\lambda) = \det(\lambda I - A) = \lambda^3 - 10\lambda^2 + 32\lambda - 32.$$

Asumimos, debido al Teorema 9.2, se cumple

$$\mathrm{e}^{tA} = x_1(t)I + x_2(t)A + x_3(t)A^2. \tag{9.5}$$

El polinomio característico produce la siguiente ecuación diferencial escalar con coeficientes constantes:

$$x^{(3)} - 10x'' + 32x' - 32x = 0.$$

La solución general es hallada en base a la ecuación característica

$$m^3 - 10m^2 + 32m - 32 = (m-2)(m-4)^2 = 0.$$

Así, la solución general toma la forma

$$x(t) = \alpha_1\,\mathrm{e}^{2t} + \alpha_2\,\mathrm{e}^{4t} + \alpha_3 t\,\mathrm{e}^{4t}\,.$$

- Para hallar $x_1(t)$ usamos las condiciones iniciales $x(0) = 1, x'(0) = 0, x''(0) = 0$. Usando esta información obtenemos

$$x_1(t) = 4\,\mathrm{e}^{2t} - 3\,\mathrm{e}^{4t} + 4t\,\mathrm{e}^{4t}\,.$$

- Para $x_2(t)$ usamos las condiciones $x(0) = 0, x'(0) = 1, x''(0) = 0$, y así lograr

$$x_2(t) = -2\,\mathrm{e}^{2t} + 2\,\mathrm{e}^{4t} - 3t\,\mathrm{e}^{4t}\,.$$

- Para $x_3(t)$ usamos las condiciones $x(0) = 0, x'(0) = 0, x''(0) = 1$, y obtenemos

$$x_3(t) = \frac{1}{4}\,\mathrm{e}^{2t} - \frac{1}{4}\,\mathrm{e}^{4t} + \frac{1}{2}t\,\mathrm{e}^{4t}\,.$$

Por último, reemplazamos estas funciones en la Fórmula (9.5) y conseguimos

$$\begin{aligned}\mathrm{e}^{tA} &= \mathrm{e}^{2t}(4 - 3\,\mathrm{e}^{2t} + 4t\,\mathrm{e}^{2t})I + \mathrm{e}^{2t}(-2 + 2\,\mathrm{e}^{2t} - 3t\,\mathrm{e}^{2t})A + \tfrac{1}{4}\,\mathrm{e}^{2t}(1 - \mathrm{e}^{2t} + 2t\,\mathrm{e}^{2t})A^2 \\ &= \begin{bmatrix} \frac{1}{4}e^{2t}\left(e^{2t}(6t-5)+9\right) & \frac{1}{4}e^{2t}\left(e^{2t}(14t-9)+9\right) & e^{4t}(3-4t) - 3e^{2t} \\ -\frac{1}{4}e^{2t}\left(e^{2t}(10t-7)+7\right) & \frac{1}{4}\left(e^{4t}(3-2t)+e^{2t}\right) & e^{4t}(3-4t) - 3e^{2t} \\ \frac{1}{4}e^{2t}\left(e^{2t}(14t-9)+9\right) & \frac{1}{4}e^{2t}\left(e^{2t}(14t-9)+9\right) & 4e^{4t}t + e^{2t} \end{bmatrix}.\end{aligned}$$

Puede verificarse que la matriz A no es diagonalizable, pero esto no es relevante para los cálculos.

¿Podemos obviar algunos cálculos en el ejemplo anterior? La respuesta es afirmativa, como lo sugiere Liz [7], en la obtención de las funciones escalares $x_i(t)$. Para ello, basta calcular la inversa de una matriz constante particular. El siguiente teorema sienta la pauta.

Teorema 9.4. *Sea A una matriz $n \times n$ constante con polinomio característico*

$$p(\lambda) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0.$$

Entonces se tiene

$$\mathrm{e}^{tA} = x_1(t)I + x_2(t)A + \cdots + x_n(t)A^{n-1},$$

donde

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix} = B_0^{-1} \begin{bmatrix} \varphi_1(t) \\ \varphi_2(t) \\ \vdots \\ \varphi_n(t) \end{bmatrix}; \tag{9.6}$$

acá B_0 es la evaluación, en $t = 0$, de la matriz

$$B_t = \begin{bmatrix} \varphi_1(t) & \varphi_1'(t) & \cdots & \varphi_1^{(n-1)}(t) \\ \varphi_2(t) & \varphi_2'(t) & \cdots & \varphi_2^{(n-1)}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_n(t) & \varphi_n'(t) & \cdots & \varphi_n^{(n-1)}(t) \end{bmatrix},$$

siempre que $S = \{\varphi_1(t), \varphi_2(t), \ldots, \varphi_n(t)\}$ sea un conjunto fundamental de soluciones para la ecuación diferencial lineal homogénea

$$x^{(n)} + c_{n-1}x^{(n-1)} + \cdots + c_1x' + c_0x = 0.$$

Prueba. Obsérvese que $p(\lambda) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0 = 0$ es la ecuación característica de la ecuación diferencial

$$x^{(n)} + c_{n-1}x^{(n-1)} + \cdots + c_1x' + c_0x = 0. \tag{9.7}$$

Debido al Teorema 9.2 tenemos

$$\mathrm{e}^{tA} = x_1(t)I + x_2(t)A + \cdots + x_n(t)A^{n-1},$$

donde $x_k(t)$ es solución de (9.7) con condiciones iniciales (9.4), para $k = 1, 2, \ldots, n$. Observemos que el conjunto $\{x_1(t), x_2(t), \ldots, x_n(t)\}$ es también un conjunto fundamental de soluciones de (9.3), ya que el wronskiano $W\big(x_1(t), \ldots, x_n(t)\big)$ toma el valor 1 en $t = 0$. Más aún, al usar el conocido teorema de unicidad de las soluciones para el problema de valores iniciales (9.3)–(9.4), tenemos

$$\varphi_k(t) = \varphi_k(0)x_1(t) + \varphi_k'(0)x_2(t) + \cdots + \varphi_k^{(n-1)}(0)x_n(t),$$

para cada $k = 1, 2, \ldots, n$, de lo cual deducimos

$$\begin{bmatrix} \varphi_1(t) \\ \varphi_2(t) \\ \vdots \\ \varphi_n(t) \end{bmatrix} = B_0 \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}.$$

Como B_0 es invertible, resulta la igualdad (9.6), lo cual completa la prueba. □

Consideremos la matriz A del Ejemplo 9.3 para apreciar la simplificación de los cálculos para la tabulación de e^{tA}. Recordemos que el polinomio característico de A es

$$p(\lambda) = \lambda^3 - 10\lambda^2 + 32\lambda - 32 = (\lambda - 2)(\lambda - 4)^2,$$

al cual asociamos la ecuación diferencial lineal

$$x^{(3)} - 10x'' + 32x' - 32x = 0.$$

Como su ecuación característica es $(\lambda - 2)(\lambda - 4)^2 = 0$, se obtiene

$$\left\{\mathrm{e}^{2t}, \mathrm{e}^{4t}, t\,\mathrm{e}^{4t}\right\}$$

como conjunto fundamental de soluciones. Con ellos formamos la matriz

$$B_t = \begin{bmatrix} \mathrm{e}^{2t} & 2\,\mathrm{e}^{2t} & 4\,\mathrm{e}^{2t} \\ \mathrm{e}^{4t} & 4\,\mathrm{e}^{4t} & 16\,\mathrm{e}^{4t} \\ t\,\mathrm{e}^{4t} & \mathrm{e}^{4t}(1+4t) & 8\,\mathrm{e}^{4t}(1+2t) \end{bmatrix},$$

de donde se calcula

$$B_0 = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 4 & 16 \\ 0 & 1 & 8 \end{bmatrix}, \quad \text{con inversa} \quad B_0^{-1} = \frac{1}{4}\begin{bmatrix} 16 & -12 & 16 \\ -8 & 8 & -12 \\ 1 & -1 & 2 \end{bmatrix}.$$

Luego, como $\mathrm{e}^{tA} = x_1(t)I + x_2(t)A + x_3(t)A^2$, tenemos por el Teorema 9.4 que se cumple

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = B_0^{-1} \begin{bmatrix} \mathrm{e}^{2t} \\ \mathrm{e}^{4t} \\ t\,\mathrm{e}^{4t} \end{bmatrix} = \begin{bmatrix} 4\,\mathrm{e}^{2t} - 3\,\mathrm{e}^{4t} + 4t\,\mathrm{e}^{4t} \\ -2\,\mathrm{e}^{2t} + 2\,\mathrm{e}^{4t} - 3t\,\mathrm{e}^{4t} \\ \frac{1}{4}\,\mathrm{e}^{2t} - \frac{1}{4}\,\mathrm{e}^{4t} + \frac{1}{2}t\,\mathrm{e}^{4t} \end{bmatrix}.$$

Al reemplazar estas funciones en la fórmula de e^{tA}, se recupera la matriz exponencial del Ejemplo 9.3.

10. Transformada de Laplace

Es usual encontrar en textos de ingeniería el método de la transformada de Laplace para resolver problemas de valores iniciales de ecuaciones lineales escalares de orden n con coeficientes constantes. Este método se generaliza a resolver el problema de valor inicial $\dot{x} = Ax$, $x(0) = x_0$. Si aplicamos transformada de Laplace obtenemos que

$$sX(s) - x(0) = AX(s) \quad \text{implica} \quad (sI - A)X(s) = x(0),$$

con lo que, al despejar, se obtiene $X(s) = (sI - A)^{-1}x(0)$. Luego de tomar transformada inversa obtenemos

$$x(t) = \mathcal{L}^{-1}\big\{(sI - A)^{-1}\big\}x(0).$$

Así, por unicidad, se logra

$$\mathrm{e}^{tA} = \mathcal{L}^{-1}\big\{(sI - A)^{-1}\big\}.$$

En aplicaciones de ingeniería, la matriz exponencial es llamada la *matriz de transición de estado*. Para ejemplos e interesantes propiedades de la matriz exponencial, ver [1].

11. Conclusiones

Se ha ilustrado varios métodos para calcular la matriz exponencial de una matriz cuadrada. La mayoría de ellos no utiliza el cálculo de autovectores (generalizados) de la matriz, lo cual ha sido un método clásico de abordar el problema en varios textos a nivel de iniciación. Si bien estos métodos pueden aplicarse a cualquier matriz, todos ellos resultan ineficaces si tratamos con matrices grandes o con entradas inexactas

(números decimales truncados o números acompañados con cierto error). Es aquí donde es preferible usar la computadora, pero como señala el trabajo de Moler-Van Loan [9], no existe un método infalible de implementación numérica. La descripción e implementación de estos métodos numéricos servirán de base para un artículo posterior.

Referencias

[1] P. J. Antsaklis & A. N. Michel; *A Linear Systems Primer*. Birkhaüser, Boston (2007).

[2] T. M. Apostol; *Some Explicit Formulas for the Exponential Matrix* e^{tA}. The American Mathematical Monthly (1969), **76**, 3:289–292.

[3] F. R. Gantmacher; *The Theory of Matrices*, Volume I. Chelsea, New York (1959).

[4] S-H Hou & W-K. Pang; *On the matrix exponential function*. International Journal of Mathematical Education in Science and Technology (2006), **37**, 1:65–70.

[5] R. B. Kirchner; *An explicit formula for* e^{At}. The American Mathematical Monthly (1967), **74**, 1200–1204.

[6] I. E. Leonard; *The Matrix Exponential*. SIAM Review (1996), **38**, 3:507–512.

[7] E. Liz; *A Note on the Matrix Exponential*. SIAM Review (1998), **40**, 3:700–702.

[8] C. D. Meyer; *Matrix Analysis and Applied Linear Algebra*. SIAM (2001).

[9] C. Moler & Ch. Van Loan; *Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later*. SIAM Review (2003), **45**, 1:3–49.

[10] E. J. Putzer; *Avoiding the Jordan Canonical Form in the Discussion of Linear Systems with Constant Coefficients*. The American Mathematical Monthly (1966), **73**, 1:2–7.

[11] S. H. Weintraub; *Jordan Canonical Form: Application to Differential Equations*. Series: Synthesis Lectures on Mathematics and Statistics. Morgan & Claypool Publishers (2008).

Abstract

We present several methods that allow the exact computation of the exponential matrix e^{tA}. Methods that include computation of eigenvectors or Laplace transform are very well-known, and they are mentioned here for completeness. We also present other methods, not well-known in the literature, that do not need the computation of eigenvectors, and are easy to introduce in a classroom, thus providing us with general formulas that can be applied to any matrix.

Keywords: Exponential matrix, diagonalizable matrix, Jordan canonical form, Schur's triangularization, functions of matrices, Lagrange-Sylvester interpolation, Putzer's spectral formula, Laplace transform.

Rubén Agapito
Sección Matemáticas
Departamento de Ciencias
Pontificia Universidad Católica del Perú
ruben.agapito@pucp.pe

Non-life Insurance Mathematics

Makoto Yamazato[1]

November, 2014

Abstract

In this work we describe the basic facts of non-life insurance and then explain risk processes. In particular, we will explain in detail the asymptotic behavior of the probability that an insurance product may end up in ruin during its lifetime. As expected, the behavior of such asymptotic probability will be highly dependent on the tail distribution of each claim.

MSC(2010): 60G07, 62P05; 97M30.

Keywords: Stochastic processes, actuarial mathematics, non-live insurance, collectice risk models, ruin probability, Cramér-Lundberg approximation.

1. *University of the Ryukyus, Japan.*

1. Introduction

Insurance mathematics, which is also called actuarial mathematics, is studied with a mathematical viewpoint in Europe and the United States. There are scientific journals which mainly deal with actuarial mathematics (e.g. Scandinavian Actuarial Journal). However, it is not a very popular topic in the Japanese mathematical community. This may also be true for Peru. There is a professional qualification for this activity called actuary. An actuary has to design, price, and control the so called insurance commodities. They play a significant role in insurance companies as they make a final decision for the design of insurance commodities. Many mathematical tools are needed in order to perform this job. The main mathematical techniques required are related to probability theory and mathematical statistics.

In order to obtain this professional qualification, one must pass an examination. The examination is divided into three parts: life insurance, non-life insurance, and annuity. Life insurance mathematics is one of the basic exams and perhaps the most important one. A life insurance contract between the insurance company and a customer is a long time bound. Hence, the interest rate for bank deposits during a long time span and the life of a person may be considered as random phenomena. Therefore, we need to know their distribution. Regarding the life of a person this is the so called life expectancy (or death probability). Some special unique symbols are used in life insurance mathematics. These symbols are unique in this world and are not commonly seen in the mathematical world. But they have been standardized worldwide, and they are also used in non-life insurance and annuity. Therefore, in order to study insurance mathematics, one should be accustomed with the use of these symbols.

On the other hand, for non-life insurance, the period of contracts are relatively short. So, the influence of interest rates is neglected. Instead, two other factors are considered: the frequency of occurrence of accidents and their size. These can be regarded, respectively, as a probability law

and a random variable with certain distribution function.

In this work, we describe the basic facts of non-life insurance and then explain risk processes, which are still an important research object in non-life insurance and applied probability. This is an application of the theory of stochastic processes. In particular, we will explain in detail the asymptotic behavior of the probability that an insurance product may end up in ruin during its lifetime. That is, given an initial investment amount set aside for payments and a premium rate per unit of time, we will discuss how to compute the probability that at some point in time the money amount of the total claims will be higher than the initial investment plus all the premium collected up to that point in time. As expected, the behavior of such asymptotic probability will be highly dependent on the tail distribution of each claim. We will call these classes light tail or heavy tailed depending on the probability of creating a large claim.

2. Non-life (damage) Insurance

There are many examples of non-life insurance: car insurance, weather related insurance, fire and earthquake insurance, aviation insurance, travel insurance, marine insurance, and so on. Let us consider for instance car insurance. Imagine a car owner. He may suffer a car accident. This accident is unpredictable, so it is a random phenomenon and, even, a rare event. Also the amount of money required for the repair of the car depends of the size of the accident and it is also unpredictable, another random phenomenon.

In this chapter, we summarize fundamental facts in non-life insurance necessary for the understanding of risk processes explained later on.

The request of insurance payment from a customer of a non-life insurance is called a **claim**. The total amount of claims in a prefixed unit time interval is the sum of the amounts of claims occurred in the time interval from all the customers. The number of claims and their

amount in each term (or fixed time interval) are not constant. So, we consider each of these quantities as random variables.

2.1 Number of claims

First, we assume the homogeneity of an objective population (group) in the sense that all contracts are equal within a population for which each individual has the same risk. Assume that each one of n contracts in a unit time interval may be the source of a claim independent of each other with given probability p. Then the probability that k claims occur in one unit of time is given by the binomial distribution ${}_nC_k p^k (1-p)^{n-k}$.

Let N be a random variable which represents the number of claims in the unit time interval. The distribution of N is the binomial distribution $B(n,p)$. If the number of contracts n is very large and the probability p of occurrence of the claim is very small, then the distribution of N can be regarded via an approximation argument as the Poisson distribution. This is explained mathematically by the following argument. Assume $\lambda = np$ and let $n \to \infty$. Then the distribution of N converges to the Poisson distribution with parameter λ since using the moment generating function formula for a binomial distribution we have

$$M_N(\theta) = E(e^{\theta N}) = (1 + (e^\theta - 1)\lambda/n)^n \to \exp(\lambda(e^\theta - 1)).$$

The right hand side is the moment generating function of the Poisson distribution with parameter λ (denoted as $Po(\lambda)$). For $N \sim Po(\lambda)$ recall that we have $E(N) = V(N) = \lambda$.

If the objective population is not homogeneous, i.e., if contracts with various risks are mixed in the population, then the variance of N is greater than its expectation ($V(N) > E(N)$). In such a case, the negative binomial distribution may fit as the distribution of N. The following is a mathematical explanation for this assertion. Let $Y_\lambda \sim Po(\lambda)$, $(\lambda > 0)$, $W \sim \Gamma(\alpha, \beta)$ (i.e., the gamma distribution with parameters α and β) and assume $\{Y_\lambda\}$ and W are independent. Let $N = Y_W$. This means the parameter of the Poisson random variable may vary according

to the value of $W(\omega)$. In this context we have

$$\begin{aligned} P(N=k) &= \int_0^\infty P(Y_\lambda = k|W=\lambda)P(W \in d\lambda) \\ &= \int_0^\infty \frac{\lambda^k}{k!}e^{-\lambda}\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}d\lambda \\ &= \frac{\beta^\alpha}{k!\Gamma(\alpha)}\int_0^\infty \lambda^{k+\alpha-1}e^{-(\beta+1)\lambda}d\lambda \\ &= \frac{\beta^\alpha\Gamma(k+\alpha)}{(\beta+1)^{k+\alpha}k!\Gamma(\alpha)} \\ &= \binom{k+\alpha-1}{k}\left(\frac{1}{\beta+1}\right)^k\left(\frac{\beta}{\beta+1}\right)^\alpha. \end{aligned}$$

In this setting, the distribution of N is the negative binomial distribution with parameter $(\alpha, \frac{\beta}{\beta+1})$, usually denoted as $NB(\alpha, \frac{\beta}{\beta+1})$. Notice that if we construe Y_λ as a stochastic process with time parameter λ (the so-called Poisson process to be defined later), then N_W is the value at time 1 of the process obtained by subordination of Y_λ by a gamma process. Another way of interpreting this structure is to say that there are various types of customers. Each type will generate its own Poisson random variable with parameter W and the density associated to each customer is given by the law of the random variable W, which itself follows the gamma distribution.

If $N \sim NB(n,p)$, then its moment generating function is given by

$$\begin{aligned} M_N(\theta) &= E(e^{\theta N}) \\ &= \sum_{k=0}^\infty e^{\theta k}\binom{k+n-1}{k}p^n(1-p)^k \\ &= \sum_{k=0}^\infty \binom{-n}{k}p^n(1-p)^k e^{\theta k} \\ &= p^n(1-(1-p)e^\theta)^{-n}. \end{aligned}$$

Hence the expectation and the variance appear as

$$E(N) = \frac{n(1-p)}{p},\ V(N) = \frac{n(1-p)}{p^2},$$

respectively. Unless $p = 1$, we always get $V(N) > E(N)$.

2.2 The total claim amount

Consider now the amount of claims. We assume the claim sizes to be positive random variables, mutually independent and identically distributed. We denote them by $X_1, X_2, \ldots$. Let S be the total amount of claims, so $S = \sum_{k=1}^{N} X_k$. Its **expectation** is

$$\begin{aligned} E(S) = E(\sum_{k=1}^{N} X_k) &= E(\sum_{k=1}^{n} X_k | N = n)P(N = n) \\ &= E(X_1)\sum_{n=0}^{\infty} nP(N = n) \\ &= E(X)E(N). \end{aligned}$$

Here X is a random variable whose law is identical with the law of all X_i. The **net premium** (i.e., the price of the insurance to be paid by each client) is given by estimating $E(N)$ and $E(S)$ from past data and then setting the net premium as $E(X) = E(S)/E(N)$. The above calculation gives $E(S|N) = NE(X)$. Let us calculate the variance. Define the **conditional variance** $V(S|N)$ by $V(S|N) = E((S - E(S|N)^2)|N)$.

Applying this equality, we obtain

$$\begin{aligned} V(S) &= E\left(E\Big((S - E(S))^2|N\Big)\right) \\ &= E\left(E\Big(\Big((S - E(S|N)) + (E(S|N) - E(S))\Big)^2|N\Big)\right) \\ &= E\Big(V(S|N)\Big) + V\Big(E(S|N)\Big). \end{aligned}$$

Here, we used the fact that $E(S|N) - E(S)$ is $\sigma(N)$-measurable. Notice that we did not need X and N to be independent at this stage in order to derive this equality. However, using the independence of X and N, we have $V(S|N) = NV(X)$ and $E(S|N) = NE(X)$. Therefore one gets

$$V(S) = E(N)V(X) + V(N)(E(X))^2.$$

In the particular case $N \sim Po(\lambda)$, then $V(N) = E(N) = \lambda$ implies

$$V(S) = \lambda((E(X))^2 + V(X)) = \lambda E(X^2).$$

The moment generating function $M_S(\theta) = E(e^{\theta S})$ of S is given by

$$\begin{aligned} M_S(\theta) &= E(E(e^{\theta S}|N)) = E(M_X(\theta)^N) \\ &= E(\exp(N \log M_X(\theta))) = M_N(\log M_X(\theta)). \end{aligned}$$

If $N \sim Po(\lambda)$, then the moment generating function of S is

$$\mathrm{M}_S(\theta) = \exp(\lambda(M_X(\theta) - 1)),$$

and the distribution function is given by

$$F_S(x) = P(S_N \leq x) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} F_X^{n*}(x),$$

where F_X is the distribution function of X and F_X^{n*} $(n \geq 1)$ is the n-fold convolution of F and $F_X^{0*}(x) = \int_{(-\infty,x]} \delta_0(dx)$. Here $\delta_0(dx)$ denotes the point mass measure at 0.

The distribution of S discussed previously corresponds to the so called **compound Poisson distribution**.

If $N \sim NB(n, p)$, we have

$$E(S) = \frac{n(1-p)}{p} E(X),\ V(S) = \frac{n(1-p)}{p} V(X) + \frac{n(1-p)}{p^2} (E(X))^2.$$

The moment generating function of S is given by

$$M_S(\theta) = p^n \Big(1 - (1-p) M_X(\theta)\Big)^{-n},$$

and the distribution function by

$$F_S(x) = \sum_{k=0}^{\infty} F_X^{k*}(x) \binom{n+k-1}{k} p^n (1-p)^k.$$

The distribution of S in this case is called the **compound negative binomial distribution**.

3. Collective Risk Models

In the previous chapter, we considered the total claim amount in a unit time interval. In this chapter, we observe the dynamical behavior of the amount of claims. We say that a family of parametrized random variables is a stochastic process regarding the parameter as a time variable. Usually the parameter runs through $[0, \infty)$ or $\{0, 1, 2, \ldots\}$.

3.1 Compound Poisson processes

A stochastic process $\{N(t); t \in [0, \infty)\}$ taking values in $\{0, 1, \ldots\}$ is called a **Poisson process with parameter** λ if it satisfies the following four conditions.
(1) $N(0) = 0$,
(2) for $0 \leq s < t$, the law of $N(t) - N(s)$ is Poisson with parameter $\lambda(t-s)$,
(3) for $0 \leq t_0 < t_1 < \cdots < t_n$, the random variables $N(t_0), N(t_1) - N(t_0), \ldots, N(t_n) - N(t_{n-1})$ are mutually independent, and
(4) $\{N(t); t \in [0, \infty)\}$ is right continuous with left limits.

Construction of a Poisson process. Let $T_1, T_2, \ldots$ be a sequence of independent random variables with common exponential distribution with parameter $\lambda > 0$ and let $S_0 = 0$, $S_n = T_1 + \cdots + T_n$ $(n \geq 1)$. Define $N_t = N(t)$ by

$$N_t = N(t) = n, \quad \text{if } S_n \leq t < S_{n+1}.$$

Obviously we have $N_0 = 0$, and N_t is by definition right continuous with left limits. Let us prove conditions (2) and (3) simultaneously. For $0 \le t_1 < t_2 < \cdots < t_n$ and $0 \le k_1 \le \cdots \le k_n$, we have

$$\begin{aligned} & P(N_{t_1} = k_1, N_{t_2} - N_{t_1} = k_2 - k_1, \ldots, N_{t_n} - N_{t_{n-1}} = k_n - k_{n-1}) \\ = \; & P(N_{t_1} = k_1, N_{t_2} = k_2, \ldots, N_{t_n} = k_n) \\ = \; & P(S_{k_1} \le t_1 < S_{k_1+1}, \cdots, S_{k_n} \le t_n < S_{k_n+1}) \\ = \; & \int \cdots \int_{D_n \cap \{t_n < s_{k_n} + u_{k_n+1}\}} \lambda^{k_n+1} e^{-\lambda(s_{k_n} + u_{k_n+1})} du_1 \cdots du_{k_n+1} \quad (1) \\ = \; & I_n, \end{aligned}$$

where $s_k = u_1 + \cdots + u_k$, for $k = 1, 2, \ldots$, and

$$D_n = \{s_1 \le \cdots \le s_{k_1} \le t_1 < s_{k_1+1} \le \cdots \le s_{k_n} \le t_n\} \subset \mathbb{R}^{k_n+1}.$$

Integrating (1) with respect to u_{k_n+1} on $[t_n - s_{k_n}, \infty)$, we obtain $I_n = \lambda^{k_n} e^{-\lambda t_n} \int_{D_n} ds_1 \cdots ds_{k_n}$. Set

$$E_\ell = \{t_{\ell-1} < s_{k_{\ell-1}+1} \le \cdots \le s_{k_\ell} \le t_\ell\}$$

for $\ell = 1, 2, \ldots, n$. Then we get $D_n = E_1 \times \cdots \times E_n$, and so

$$\int_{E_\ell} ds_{k_{\ell-1}} \cdots ds_{k_\ell} = \frac{(t_\ell - t_{\ell-1})^{k_\ell - k_{\ell-1}}}{(k_\ell - k_{\ell-1})!}.$$

Hence we have

$$I_n = \frac{(\lambda t_1)^{k_1}}{k_1!} e^{-\lambda t_1} \prod_{i=2}^{n} \frac{\{\lambda(t_i - t_{i-1})\}^{k_i - k_{i-1}}}{(k_i - k_{i-1})!} e^{-\lambda(t_i - t_{i-1})}.$$

Letting $n = 2$ we obtain

$$\begin{aligned} P(N_t - N_s = k) &= \sum_{j=0}^{\infty} P(N_s = j, N_t - N_s = k) \\ &= \sum_{j=0}^{\infty} \frac{(\lambda s)^j}{j!} e^{-\lambda s} \frac{\{\lambda(t-s)\}^k}{k!} e^{-\lambda(t-s)} \\ &= \frac{\{\lambda(t-s)\}^k}{k!} e^{-\lambda(t-s)}. \end{aligned}$$

Plugging this last relation into the first one we conclude the equality

$$\begin{aligned} & P(N_{t_1} = k_1, N_{t_2} - N_{t_1} = k_2 - k_1, \dots, N_{t_n} - N_{t_{n-1}} = k_n - k_{n-1}) \\ = \quad & P(N_{t_1} = k_1) \prod_{i=2}^{n} P(N_{t_i} - N_{t_{i-1}} = k_i - k_{i-1}). \end{aligned}$$

Next we define what a Lévy process is. A stochastic process $\{X_t; t \in [0,\infty)\}$ on $\mathbb{R}^d$ is called a **Lévy process** if it satisfies the following four conditions.
(1) $X_0 = 0$,
(2) for any $n \geq 1$ and $0 \leq t_0 < t_1 < \cdots < t_n$, the random variables $X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}}$ are mutually independent,
(3) for $0 \leq s < t$, the distribution of $X_t - X_s$ is identical to the distribution of X_{t-s}, and
(4) $\{X_t; t \in [0,\infty)\}$ is right continuous with left limits.

In the sequel $(a \wedge b)$ represents the minimum between $a, b \in \mathbb{R}$.

Theorem 3.1. *If $\{X_t\}$ is a Lévy process on $\mathbb{R}^d$, then its characteristic function $E(e^{i<\theta, X_t>})$, with $\theta \in \mathbb{R}^d$, is given by*

$$\exp[t\{i\langle\theta,\gamma\rangle - \frac{|\sigma\theta|^2}{2} + \int_{\mathbb{R}^d\setminus\{0\}} (e^{i<\theta,x>} - 1 - i\langle\theta,x\rangle 1_{\{|x|\leq 1\}})\nu(dx)\}],$$

here γ is in $\mathbb{R}^d$ and σ is a $d \times d$-symmetric nonnegative definite matrix, while ν is a nonnegative measure on $\mathbb{R}^d\setminus\{0\}$ satisfying $\int(|x|^2 \wedge 1)\nu(dx) < \infty$.

Proof. We refer to Sato [7] for the proof. □

A Lévy process $\{X_t\}$ on $\mathbb{R}^d$ is called a **compound Poisson process** if its characteristic function is represented as

$$\exp\{t \int_{\mathbb{R}^d\setminus\{0\}} e^{i<\theta,x>} - 1\, \nu(dx)\}.$$

Here $\nu(dx)$ is a measure on $\mathbb{R}^d\backslash\{0\}$ that satisfies $\int_{\mathbb{R}^d\backslash\{0\}}\nu(dx)<\infty$.

Construction of a compound Poisson processes. Let F be a probability measure on $\mathbb{R}^d\backslash\{0\}$ and $X_1, X_2, \cdots$ be a sequence of $\mathbb{R}^d$-valued independent random variables with common distribution F. Let $\{N_t\}$ be a Poisson process with parameter $\lambda>0$, independent from $X_1, X_2, \cdots$. Then, the stochastic process defined by

$$Y_t=\begin{cases}0 & \text{if } N_t=0,\\ X_1+\cdots+X_{N_t} & \text{otherwise}\end{cases}$$

is a compound Poisson process with $\nu(dx)=\lambda F(dx)$.

Let us proof this fact. For times $0=t_0\le t_1<t_2\ldots<t_n$ and sets $B_1, B_2, \ldots, B_n\in\mathcal{B}(\mathbb{R}^d)$ we have (the sums are always taken along all possible $0=m_0\le m_1\le m_2\le\cdots\le m_n$)

$$\begin{aligned}
&P(Y_{t_1}\in B_1, Y_{t_2}-Y_{t_1}\in B_2,\ldots,Y_{t_n}-Y_{t_{n-1}}\in B_n)=\\
&=\sum\prod_{i=1}^{n}P(X_{m_{i-1}+1}+\cdots+X_{m_i}\in B_i)P(N_{t_i}-N_{t_{i-1}}=m_i-m_{i-1})\\
&=\sum\prod_{i=1}^{n}P(X_1+\cdots+X_{m_i-m_{i-1}}\in B_i)P(N_{t_i-t_{i-1}}=m_i-m_{i-1})\\
&=\sum\prod_{i=1}^{n}P(Y_{m_i-m_{i-1}}\in B_i, N_{t_i-t_{i-1}}=m_i-m_{i-1})\\
&=\prod_{i=1}^{n}P(Y_{t_i-t_{i-1}}\in B_i).
\end{aligned}$$

On the other hand, if $t\ge s$, from

$$\begin{aligned}
P(Y_t-Y_s\in B) &=\sum_{0\le m\le n}P(X_{m+1}+\cdots+X_n\in B, N_s=m, N_t=n)\\
&=\sum_{0\le m\le n}P(X_1+\cdots+X_{n-m}\in B, N_s=m, N_t=n)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{0\le m\le n} P(X_1+\cdots+X_{n-m}\in B, N_{t-s}=n-m)\times \\
&\qquad\qquad \times P(N_s=m) \\
&= \sum_{0\le m\le n} P(Y_{t-s}\in B, N_{t-s}=n-m)P(N_s=m) \\
&= P(Y_{t-s}\in B)
\end{aligned}
$$

we get $Y_t - Y_s \sim Y_{t-s}$.

Also, by the very definition of Y we get

$$
\begin{aligned}
P(Y_t\in A) &= \delta_0(A)+\sum_{n\ge 1} P(X_1+\cdots+X_n\in A)P(N(t)=n) \\
&= \sum_{n\ge 0} F^{n*}(A)\frac{(\lambda t)^n}{n!}e^{-\lambda t}.
\end{aligned}
$$

Hence we obtain

$$
\begin{aligned}
E(e^{i<\theta,Y_t>}) &= \sum_{n\ge 0}\int e^{i<\theta,y>}F^{n*}(dy)\frac{(\lambda t)^n}{n!}e^{-\lambda t} \\
&= \sum_{n\ge 0}(\int e^{i<\theta,y>}F(dy))^n\frac{(\lambda t)^n}{n!}e^{-\lambda t} \\
&= \exp\{t\int_{R^d\setminus\{0\}}(e^{i<\theta,x>}-1)\nu(dx)\}.
\end{aligned}
$$

3.2 Risk processes

Let Y_t be a compound Poisson process on $[0,\infty)$ with characteristic function

$$E(e^{i\theta Y_t}) = \exp\{\lambda t\int_{(0,\infty)}(e^{i\theta x}-1)\,F(dx)\}.$$

We call $R(t) = ct - Y_t$ the **risk process** and $U(t) = u + R(t)$ the **surplus reserve process**. We assume here that the expectation $\mu =$

$\int_{(0,\infty)} xF(dx) = \int_0^\infty 1 - F(x)\,dx$ of F is finite. We call the parameter $\rho = \frac{c}{\lambda\mu} - 1$ **the safety loading**. We always assume $\rho > 0$.

Let τ_u be the **hitting time** of $(-\infty, 0)$ for $U(t)$; that is,

$$\tau_u = \inf\{t \geq 0; U(t) \in (-\infty, 0)\}.$$

This hitting time is called **ruin time** and $\mathcal{E}(u) = P(\tau_u < \infty)$ the **ruin probability** in insurance mathematics. Here c represents the rate of premiums that the insurer receives per unit of time and u is the initial capital of the insurance company.

Lemma 3.2. *For $A, B, C \in \mathcal{B}(\mathbb{R})$, we have*

$$P(R(T_1) \in A, R(T_1 + t) \in B, T_1 \in C) = \int_C \int_A P(z + R(t) \in B)\, P(cs - X_1 \in dz)\, P(T_1 \in ds).$$

Proof. For $A, B, C \in \mathcal{B}(\mathbb{R})$, we have

$$\begin{aligned}
& P(R(T_1) \in A, R(T_1 + t) \in B, T_1 \in C) = \\
= & \sum_{m=1}^{\infty} \int_C P(cs - X_1 \in A, N_{s+t} - N_s = m - 1, R(T_1 + t) \in B | T_1 = s) \\
& \qquad \times P(T_1 \in ds) \\
= & \sum_{m=1}^{\infty} \int_C P(cs - X_1 \in A, N_s = 1, N_{s+t} - N_s = m - 1, \\
& \qquad c(s+t) - (X_1 + \cdots + X_m) \in B | T_1 = s)\, P(T_1 \in ds) \\
= & \sum_{m=1}^{\infty} \int_C P(ct - X_1 \in A, T_2 + \cdots + T_m \leq t < T_2 + \cdots + T_{m+1}, \\
& \qquad c(s+t) - (X_1 + \cdots + X_m) \in B | T_1 = s)\, P(T_1 \in ds) \\
= & \sum_{m=1}^{\infty} \int_C P(cs - X_1 \in A, c(s+t) - (X_1 + \cdots + X_m) \in B) \\
& \qquad \times P(T_2 + \cdots + T_m \leq t < T_2 + \cdots + T_{m+1})\, P(T_1 \in ds)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{m=1}^{\infty} P(N_t = m-1) \times \\
&\quad \times \int_C P(cs - X_1 \in A, cs - X_1 + ct - (X_2 \cdots + X_m) \in B)\, P(T_1 \in ds) \\
&= \sum_{m=1}^{\infty} P(N_t = m-1) \times \\
&\quad \times \int_C \int_A P(z + ct - (X_2 \cdots + X_m) \in B)\, P(cs - X_1 \in dz) P(T_1 \in ds) \\
&= \int_C \int_A P(z + R(t) \in B)\, P(cs - X_1 \in dz)\, P(T_1 \in ds).
\end{aligned}
$$

□

The following theorem exhibits a closed formula for the ruin probability based on the claim distribution.

Theorem 3.3. *Let $h(y) = \frac{1-F(y)}{\mu}$ and $H(x) = \int_0^x h(y)dy$. Then H is a distribution function and the following equality*

$$\mathcal{E}(u) = 1 - \frac{\rho}{1+\rho} \sum_{n=0}^{\infty} \frac{H^{n*}(u)}{(1+\rho)^n}$$

holds.

Proof. Let $\varphi(u) = 1 - \mathcal{E}(u)$ be the chance of never get ruined. As the ruin can never happen before the first event, we have

$$
\begin{aligned}
\varphi(u) &= P(U(t) \geq 0, \text{for all } t \geq 0) \\
&= P(u + R(T_1) \geq 0, u + R(T_1 + t) \geq 0, \text{for all } t \geq 0) \\
&= \int_0^{\infty} \int_{(0,u+cs]} P(u + cs - z + R(t) \geq 0, \text{for all } t \geq 0) \times \\
&\quad \times P(X_1 \in dz)\, P(T_1 \in ds) \\
&= \int_0^{\infty} \lambda e^{-\lambda s} \{ \int_{(0,u+cs]} \varphi(u + cs - z) F(dz) \} ds.
\end{aligned}
$$

With the change of variable $t = u + cs$, we arrive at

$$\varphi(u) = \frac{\lambda}{c} e^{\lambda u/c} \int_u^\infty e^{-\lambda t/c} \{ \int_{(0,t]} \varphi(t-z) F(dz) \} dt.$$

Hence $\varphi(u)$ is continuous, right differentiable, and satisfies

$$\varphi'(u) = \frac{\lambda}{c} \varphi(u) - \frac{\lambda}{c} \{ \int_{(0,u]} \varphi(u-z) F(dz) \}. \tag{2}$$

Noting that by Fubini's theorem we have

$$\begin{aligned} \int_0^u \Big(\int_{(0,v]} \varphi(v-x) F(dx) \Big) dv &= \int_{(0,u]} (\int_0^{u-x} \varphi(v) dv) F(dx) \\ &= \int_0^u \varphi(v) (\int_{(0,u-v]} F(dx)) dv \\ &= \int_0^u \varphi(u-v) F(v) dv, \end{aligned} \tag{3}$$

one then integrates (2) to obtain

$$\begin{aligned} \varphi(u) &= \varphi(0) + \frac{\lambda}{c} \left(\int_0^u \varphi(v) dv - \int_0^u (\int_{(0,v]} \varphi(v-x) F(dx)) dv \right) \\ &= \varphi(0) + \frac{\lambda}{c} \int_0^u \varphi(u-z)[1-F(z)] dz. \\ &= \varphi(0) + \frac{1}{1+\rho} \int_0^u \varphi(u-z) H(dz). \end{aligned} \tag{4}$$

Letting $u \to \infty$, by monotone convergence we get

$$\varphi(\infty) = \varphi(0) + \frac{\lambda\mu}{c} \varphi(\infty), \tag{5}$$

where $\mu = \int_0^\infty 1 - F(z)\, dz$.

On the other hand, let

$$\Omega_0 = \{ \omega \in \Omega : \lim_{t\to\infty} \frac{R_t(\omega)}{t} = c - \lambda\mu \}.$$

Then the strong law of large numbers implies $\lim_{t\to\infty} \frac{N_t(\omega)}{t} = \lambda > 0$ and $\lim_{t\to\infty} \frac{1}{N_t(\omega)} \sum_{i=1}^{N_t(\omega)} X_i(\omega) = \mu$ for almost all ω. Therefore $P(\Omega_0) = 1$.

Since $c > \lambda\mu$, for $\omega \in \Omega_0$, there is $u > 0$ such that for all $t \geq 0$ we have $R_t(\omega) > -u$. Hence we get

$$\begin{aligned} \varphi(\infty) &= \lim_{u\to\infty} \varphi(u) \\ &= \lim_{u\to\infty} P(\cap_{t\geq 0}\{R_t > -u\}) \\ &= \lim_{n\to\infty} P(\cap_{t\in\mathbb{Q}_+}\{R_t > -n\}) \\ &= P(\cup_{n\geq 1} \cap_{t\in\mathbb{Q}_+} \{R_t > -n\}) = 1. \end{aligned}$$

By this equality and (5), we have

$$\varphi(0) = \frac{\rho}{1+\rho} \quad \text{and} \quad \mathcal{E}(0) = \frac{1}{1+\rho}.$$

Set $\widehat{H}(\theta) = \int_0^\infty e^{-\theta x} h(x)dx$ and $\widehat{\varphi}(\theta) = \int_{[0,\infty)} e^{-\theta x}\varphi(dx)$. Then by (4) we get

$$\widehat{\varphi}(\theta) = \frac{\rho}{1+\rho} + \frac{1}{1+\rho}\widehat{\varphi}(\theta)\widehat{H}(\theta).$$

So at the end we conclude

$$\widehat{\varphi}(\theta) = \frac{\frac{\rho}{1+\rho}}{1 - \frac{1}{1+\rho}\widehat{H}(\theta)}.$$

□

Example 3.4. Let $F(x) = 1 - e^{-x/\mu}$. Then $H(x) = 1 - e^{-x/\mu}$ and hence by the above theorem, $\mathcal{E}(x) = \frac{1}{1+\rho} e^{-\rho x/(1+\rho)\mu}$.

Remark 3.5. Let $Y_1, Y_2, \ldots$ be a sequence of independent random variables with distribution function H, and M be a random variable independent of $Y_1, Y_2, \ldots$ which have geometric distribution with parameter $\rho(1+\rho)^{-1}$, that is,

$$P(M = n) = \frac{\rho}{1+\rho}(1+\rho)^{-n}, \ n = 0, 1, \cdots.$$

Then we have

$$\mathcal{E}(u) = P(Y_1 + Y_2 + \cdots + Y_M > u).$$

Theorem 3.6. *For* $G(u,y) = P(U(\tau_u) \in [-y,0), \tau_u < \infty)$ *we have* $G(0,y) = \frac{\lambda}{c}\int_0^y (1 - F(u)du$.

Proof. In the case of our concern, we must have $\tau_u = \sum_{k=1}^m T_k$ for some $m \geq 1$. As in Lemma 3.2, we have then

$$\begin{aligned} G(u,y) &= P(U(T_1) \in [-y,0)) \\ &\quad + \sum_{m=2}^{\infty} P(U(T_1) \geq 0, \ldots, U(\sum_{k=1}^{m-1} T_k) \geq 0, U(\sum_{k=1}^{m} T_k) \in [-y,0)) \\ &= \int\int_{-y \leq u+ct-x<0} \lambda e^{-\lambda t} F(dx)dt \\ &\quad + \int\int_{0 \leq u+ct-x<u+ct} G(u+ct-x,y)\lambda e^{-\lambda t} F(dx)dt. \end{aligned}$$

However, the first integral equals

$$\begin{aligned} &= \int_0^{\infty} \{F(u+ct+y) - F(u+ct)\}\lambda e^{-\lambda t} dt \\ &= \int_u^{\infty} \{F(s+y) - F(s)\}\frac{\lambda}{c} e^{-\lambda(s-u)/c} ds \\ &= \frac{\lambda}{c} e^{\lambda u/c} \int_u^{\infty} \{F(s+y) - F(s)\} e^{-\lambda s/c} ds, \end{aligned}$$

while the second is

$$\begin{aligned} &= \int_0^{\infty} \Big(\int_{(0,u+ct]} G(u+ct-x,y)F(dx) \Big) \lambda e^{-\lambda t} dt \\ &= \int_u^{\infty} \Big(\int_{(0,s]} G(s-x,y)F(dx) \Big) \frac{\lambda}{c} e^{-\lambda(s-u)/c} ds. \end{aligned}$$

From the above calculations we conclude that G is right differentiable in u and satisfies

$$\frac{\partial}{\partial u}G(u,y) = \frac{\lambda}{c}[G(u,y) - \{F(u+y) - F(u)\} - \int_{(0,u]} G(u-x,y)F(dx)].$$

Integrating from 0 to M with respect to u, we have

$$\begin{aligned} G(M,y) - G(0,y) &= -\frac{\lambda}{c}\Big[\int_0^M \{F(u+y) - F(u)\}du \\ &\quad + \int_0^M \{G(u,y) - \int_{(0,u]} G(u-x,y)F(dx)\}du\Big]. \end{aligned}$$

By (3) we get

$$\begin{aligned} &\int_0^M \{G(u,y) - \int_0^u G(u-x,y)F(dx)\}du = \\ &\qquad = \int_0^M G(M-x,y)(1-F(x))dx \to 0 \quad (M \to \infty). \end{aligned}$$

Here we have applied the equality $G(\infty, y) = 0$ and the Lebesgue's dominated convergence theorem. Hence we conclude

$$\begin{aligned} -G(0,y) &= G(\infty,y) - G(0,y) \\ &= -\frac{\lambda}{c}\int_0^\infty \{F(u+y) - F(u)\}du \\ &= -\frac{\lambda}{c}\int_0^\infty \{(F(u+y) - 1) + (1 - F(u))\}du \\ &= -\frac{\lambda}{c}\int_0^y (1 - F(u))du. \end{aligned}$$

□

Let F, G be two distribution functions. The number $L(F,G) = \inf\{h > 0 : F(x-h) - h \le G(x) \le F(x+h) + h, \text{ for all } x \in \mathbb{R}\}$ is called **the Lévy's distance from F to G.**

Problem 3.7. Show that L satisfies the axioms of distance in a metric space.

Lemma 3.8. *Let F_n, F be distribution functions. For every continuity point of F we have $\lim F_n(x) = F(x)$ if and only if $\lim L(F_n, F) = 0$.*

Proof. We first deal with the reverse implication. If $h \geq L(F_n, F)$, then for each x we have

$$F(x-h) - h \leq F_n(x) \leq F(x+h) + h.$$

As the inequality

$$F(x-h) - h \leq F(x) \leq F(x+h) + h$$

always holds, we get $|F(x) - F_n(x)| \leq F(x+h) - F(x-h) + 2h$. Hence, if x is a continuity point of $F(x)$, we have the limit $\lim_{n\to\infty} F_n(x) = F(x)$.

Conversely, choose $0 < \epsilon < 1$ arbitrary. As any increasing right continuous function has at most a countable number of discontinuities, we can choose continuity points $x_0 < x_1 < \ldots, x_k$ of F so that $F(x_0) < \epsilon$, $F(x_k) > 1 - \epsilon$, and $x_i - x_{i-1} < \epsilon$, $i = 1, \ldots, k$. Choose n_0 so that $|F(x_i) - F_n(x_i)| < \epsilon$ for all $0 \leq i \leq k$ and $n \geq n_0$. Then for x with $x_{i-1} < x < x_i$ we get

$$\begin{aligned} F_n(x-\epsilon) - \epsilon \leq F_n(x_{i-1}) - \epsilon &< F(x_{i-1}) \\ &\leq F(x) \\ &\leq F(x_i) \leq F_n(x_i) + \epsilon \leq F_n(x+\epsilon) + \epsilon, \end{aligned}$$

and for $x < x_0$,

$$F_n(x-2\epsilon) - 2\epsilon \leq F_n(x_0) - \epsilon - F(x_0) \leq 0 \leq F(x) < \epsilon \leq \epsilon + F_n(x+\epsilon).$$

For $x > x_k$, we have the similar estimate, thus we obtain $L(F_n, F) \leq 2\epsilon$. □

Lemma 3.9. *Let F, G be distribution functions. If G has a bounded density g, then we have*

$$\sup_x |F(x) - G(x)| \leq (1 + \sup_x g(x)) L(F, G).$$

Proof. Set $g^* = \sup_x g(x)$. Let $h > 0$ be so that $G(x-h) - h \le F(x) \le G(x+h) + h$ holds for any $x > 0$. Then as we have

$$G(x) - h(g^*+1) \le G(x-h) - h \le F(x) \le G(x+h) + h \le G(x) + h(g^*+1),$$

we obtain $|F(x) - G(x)| \le h(1+g^*)$. Finally, take the supremum with respect to x and then the infimum with respect to h in order to stablish the result. □

Theorem 3.10 (Rényi)**.** *Let $Y_1, Y_2, \dots$ be a sequence of nonnegative independent and identically distributed random variables. Let M_ϵ be a geometric random variable with parameter ϵ independent of $\{Y_n\}_{n\ge1}$. Set $S_\epsilon = Y_1 + \cdots + Y_{M_\epsilon}$ and let $\alpha = EY_1$ and $F_\epsilon(x) = P(\epsilon\alpha^{-1}S_\epsilon \le x)$. We have then*

$$\lim_{\epsilon\to0} \sup_{x\ge0} |F_\epsilon(x) - 1 + e^{-x}| = 0.$$

Proof. Conditioning with respect to M_ϵ and using the independent and identically distributed property of the sequence Y_i, one obtains

$$\begin{aligned} E(\exp(i\theta\epsilon\alpha^{-1}S_\epsilon)) &= \sum_{n=0}^{\infty} E[\exp\{i\theta\epsilon\alpha^{-1}(Y_1+\cdots+Y_n)\}]P(M_\epsilon = n) \\ &= \sum_{n=0}^{\infty} (E\exp(i\theta\epsilon\alpha^{-1}Y_1))^n \epsilon(1-\epsilon)^n \\ &= \frac{\epsilon}{1-(1-\epsilon)E\exp(i\theta\epsilon\alpha^{-1}Y_1)} \to \frac{1}{1-i\theta} \ (\epsilon\to0). \end{aligned}$$

Note that for the last equality we have used L'Hopital's rule. Then we have

$$\lim_{\epsilon\to0} |F_\epsilon(x) - 1 + e^{-x}| = 0$$

for all $x \ge 0$ by Lemma 3.8 (see also Lemma 4.17). Finally, by Lemmas 3.8 and 3.9, one obtains the claimed result. □

Theorem 3.11. *If $E(X_1^2) < \infty$, then we have*

$$\lim_{\rho\to 0}\sup_{u\geq 0}\Big|\mathcal{E}(u) - \exp\{\frac{-2\rho\mu u}{(1+\rho)E(X_1^2)}\}\Big| = 0.$$

Proof. Let $Y_1, Y_2, \cdots$ be nonnegative independent identically distributed random variables with density function $\frac{1-F(x)}{\mu}$ and let M be a geometric random variable with parameter $\frac{\rho}{1+\rho}$ independent of $Y_1, Y_2, \ldots$, that is, that behave like

$$P(M = n) = \frac{\rho}{1+\rho}(1+\rho)^{-n},\ n = 0, 1, \cdots.$$

By Remark 3.5, we have then

$$\mathcal{E}(u) = P(Y_1 + \cdots + Y_M > u).$$

Set $\epsilon = \frac{\rho}{1+\rho}$, $S_\epsilon = Y_1 + \cdots + Y_M$ and $x = \frac{\rho}{1+\rho}\frac{u}{E(Y_1)}$, and then apply Rényi's theorem. So we have

$$\begin{aligned} &\lim_{\rho\to 0}\sup_{u\geq 0}\Big|\mathcal{E}(u) - \exp\{\frac{-\rho u}{(1+\rho)EY_1}\}\Big| \\ = &\lim_{\epsilon\to 0}\sup_{x}\Big|P(\frac{\epsilon S_\epsilon}{E(Y_1)} > x) - e^{-x}\Big| = 0. \end{aligned}$$

Due to integration by parts we have $E(Y_1) = \mu^{-1}\int_0^\infty y(1-F(y))dy = \frac{1}{2\mu}E(X_1^2)$. The conclusion follows. □

Example 3.12. If the distribution of the random variable X_1 is an exponential distribution with mean μ, then because of $E(X_1^2) = 2\mu^2$ and Example 3.4 we obtain

$$|\mathcal{E}(u) - \exp\{\frac{-2\rho\mu u}{(1+\rho)E(X_1^2)}\}| = \frac{\rho}{1+\rho}\exp\{\frac{-\rho u}{(1+\rho)\mu}\}.$$

Hence we get

$$\lim_{\rho\to 0}\sup_{u\geq 0}|\mathcal{E}(u) - \exp\{\frac{-2\rho\mu u}{(1+\rho)E(X_1^2)}\}| = \lim_{\rho\to 0}\frac{\rho}{1+\rho} = 0,$$

and the asymptotic behavior of ruin probability as $\rho \to 0$ is compatible with the conclusion of Theorem 3.11.

4. Asymptotics of Ruin Probabilities (light tail claims)

In the rest of this article we will describe the asymptotic behavior of the ruin probability as the initial surplus u tends to infinity. First we will discuss the Cramér-Lundberg approximation which is treated in standard actuarial textbooks. Roughly speaking, this is related to an exponential decay of the tail of the claim distribution.

4.1 Cramér-Lundberg approximation

A distribution function $F(x)$ satisfies the **Cramér-Lundberg condition** if there is an $r^* > 0$ such that

$$\frac{\lambda}{c}\int_0^\infty e^{r^*x}[1-F(x)]dx = 1.$$

The parameter r^* is called the **Cramér-Lundberg exponent** or **adjustment coefficient**. In actuarial textbooks this exponent is usually denoted by R. Here, in order to avoid confusion, we use r^*. The above equation is rewritten using the integration by parts formula as

$$\int_{[0,\infty)} e^{r^*x}F(dx) = 1 + \frac{cr^*}{\lambda}.$$

Theorem 4.1. *We have*

$$\mathcal{E}(u) \le E(\exp[-r^*U(\tau_u)] : \tau_u < \infty) = e^{-r^*u}.$$

Proof. Since $U(\tau_u) \le 0$, the first inequality is obvious. Define

$$Z(t) = \exp[-r^*\{R(t)+u\}]$$

and $\mathcal{F}_t = \sigma\{Z(s); s \le t\}$. Then $Z(t)$ is an $\{\mathcal{F}_t\}$-martingale. In fact, by the Cramér-Lundberg condition and the independent increment property

of the risk process, we have

$$\begin{aligned} E(Z(t+s)|\mathcal{F}_t) &= e^{-r^*[R(t)+u]}E(e^{-r^*R(s)}) \\ &= Z(t)\exp[\lambda s(E(e^{r^*X_1})-1)]e^{-r^*cs} \\ &= Z(t). \end{aligned}$$

Therefore by the optional sampling theorem, we obtain

$$E(Z(\tau_u \wedge t)|\mathcal{F}_0) = Z(0) = e^{-r^*u}.$$

On the other hand,

$$\begin{aligned} E(Z(\tau_u \wedge t)|\mathcal{F}_0) &= E(Z(\tau_u \wedge t)) \\ &= E(Z(\tau_u) : \tau_u \leq t) + E(Z(t) : \tau_u > t) \end{aligned}$$

holds. The first term of the right hand side of the above equality tends to $E(Z(\tau_u) : \tau_u < \infty)$ as $t \to \infty$. By the strong law of large numbers we get $\lim_{t\to\infty} Z(t) = 0$ a.s. and $R(t,\omega) + u \geq 0$ for $\omega \in \{\tau_u(\omega) > t\}$; hence $0 \leq Z(t)1_{\{\tau_u > t\}} \leq 1$. Now, by the bounded convergence theorem, we get

$$\begin{aligned} \lim_{t\to\infty} E(Z(t)1_{\{\tau_u>t\}}) &= E(\lim_{t\to\infty} Z(t)1_{\{\tau_u>t\}}) \\ &= 0. \end{aligned}$$

Therefore, we must have

$$E(Z(\tau_u) : \tau_u < \infty) = e^{-r^*u}.$$

□

Theorem 4.2 (The law of large numbers)**.** *If $R(t)$ is the risk process, then we have*

$$P\Big(\lim_{t\to\infty} \frac{R(t)}{t} = E(R(1))\Big) = 1.$$

Proof. Note that $R(n) = \sum_{i=0}^{n-1}(R(i+1) - R(i)), n = 1, 2, \ldots,$ are the sums of independent and identically distributed random variables and $E(R(n)) = nE(R(1))$ is finite. By the usual strong law of large numbers, we can thus conclude

$$P(\lim_{n\to\infty} \frac{R(n)}{n} = E(R(1))) = 1.$$

Set $V_n = \sup_{n<t\le n+1} |R(t) - R(n)|$. Then $\{V_n\}$ is a sequence of independent and identically distributed random variables. We show that $E(V_0) < \infty$. Let $S_j = R(j2^{-n})$, $M_0 = 0$, and $M_k = \max_{1\le j\le k} |S_j|$, $k = 1, 2, \ldots, m = 2^n$. For $a, b > 0$, write $A_k = \{M_{k-1} \le a + b < |S_k|\}$. Then we have $\{M_m > a + b\} = \cup_{k=1}^m A_k$ and $A_j \cap A_k = \emptyset \quad (j \ne k)$. So we get

$$\begin{aligned} P(|S_m| > a) &\ge \sum_{k=1}^m P(A_k \cap \{|S_m| > a\}) \\ &\ge \sum_{k=1}^m P(A_k \cap \{|S_m - S_k| \le b\}) \\ &= \sum_{k=1}^m P(A_k)P(|S_m - S_k| \le b) \\ &\ge P(M_m > a + b) \min_{1\le k\le m} P(|S_m - S_k| \le b). \end{aligned}$$

As $P(M_m \le b/2) \le P(|S_m - S_k| \le b)$ is verified for $k = 1, \cdots, m$, we get

$$P(|S_m| > a) \ge P(M_m > a + b)P(M_m \le b/2).$$

We obtain $P(|R(1)| > a) \ge P(V_0 > a + b)P(V_0 \le b/2)$ letting $m \to \infty$. From

$$\int_{(0,\infty)} P(|R(1)| > x)dx = \int_{(0,\infty)} xP(|R(1)| \in dx) = E(|R(1)|)$$

and

$$E(V_0 - b : V_0 > b) \le \int_{(0,\infty)} P(V_0 > x + b)dx \le E(|R(1)|)/P(V_0 \le b/2)$$

(with b big enough) we can conclude then $E(V_0) < \infty$. The strong law of large numbers implies so $\lim_{n\to\infty} \frac{V_n}{n} = 0$. For $n \le t < n+1$, we have

$$\left|\frac{R(t)}{t} - E(R(1))\right| \le \left|\frac{R(n)}{n} - E(R(1))\right| + \frac{|R(t) - R(n)|}{n} + \frac{|R(n)|}{n^2},$$

and hence the conclusion. □

Define $K(x) = \sum_{n=0}^{\infty} F^{n*}(x)$ for a distribution function F on $(0,\infty)$. Note that for any $x > 0$ there is ℓ such that $F^{\ell *}(x) < 1$ We also have

$$F^{n*}(x) = \int_0^x F(x-y)dF^{(n-1)*}(y) \le F(x)F^{(n-1)*}(x) \le F^n(x) \quad (6)$$

for $n \ge 2$. Then, we obtain

$$\sum_{n=0}^{\infty} F^{n*}(x) \le \ell \sum_{k=0}^{\infty} (F^{\ell *}(x))^k < \infty.$$

Theorem 4.3 (Renewal equation)**.** *Let F be a distribution function on $(0,\infty)$ and $y : [0,\infty) \to \mathbb{R}$ be a bounded Borel measurable function. Then the bounded measurable solution of the renewal equation*

$$z(x) = y(x) + \int_0^x z(x-u)F(du) \quad (7)$$

which vanishes on $x < 0$ is $z(x) = \int_0^x y(x-u)K(du)$.

Proof. Let z_1, z_2 be two bounded measurable solutions of (7). Then $z = z_1 - z_2$ satisfies $z = z * F$. By this equality, $z = z * F^{n*}$ holds for any $n \ge 1$. Letting $n \to \infty$, we have $z(x) \le \left(\sup_{u\ge 0} z(u)\right) F^{n*}(x) \to 0$. Finally, the fact that $\int_0^x y(x-u)K(du)$ satisfies the equation (7) is straightforward. □

A distribution function F on $\mathbb{R}$ is called **arithmetic** if there is $\lambda > 0$ such that the points of increase of F are contained in the set $\{n\lambda : n = 0, \pm 1, \pm 2, \ldots\}$. Otherwise, F is **non-arithmetic**.

Lemma 4.4. *Let F be a distribution on $(0,\infty)$ and Σ a set formed by the points of increase of $F, F^{2*}, F^{3*}, \ldots$. If F is non-arithmetic, then for given $\epsilon > 0$ and x sufficiently large the interval $(x, x+\epsilon)$ contains points of Σ.*

Proof. Let $0 < a < b$ be two points in the set Σ and put $h = b - a$. Let $I_n = (na, nb) = (na, na + nh)$. If $n > a/h$, $(na, (n+1)a]$ is contained in I_n and hence every point $x > x_0 = a^2/h$ belongs to at least one among the intervals $I_1, I_2, \ldots$. The $n+1$ points $na + kh$, $k = 0, \ldots, n$, belong to Σ, and they partition I_n into n subintervals of length h. Thus every point $x > x_0$ is at a distance $\leq h/2$ from a point in Σ. Suppose that there exists $\delta > 0$ such that $h \geq \delta$ is satisfied for all possible choices. It follows that the points $na + kh$ exhaust all points of $\Sigma \cap I_n$. Since $(n+1)a \in \Sigma$ and $(n+1)a < nb$ are satisfied, there is $1 \leq k \leq n$ for which we have $na + a = na + kh$ and hence $a = kh$, and thus all points of $\Sigma \cap I_n$ are multiples of h. Now let c be an arbitrary point of increase of F. For n sufficiently large the interval I_n contains a point of the form $kh + c$, and, as this belongs to Σ, it follows that c is a multiple of h. This shows that the distribution F is arithmetic. This is a contradiction. Hence for each $\epsilon > 0$ it is possible to choose a, b in order to have $h < \epsilon$. □

Proposition 4.5. *Let F be a non-arithmetic distribution function on $(0,\infty)$. If the uniformly continuous solution of*

$$z(x) = \int_{\mathbb{R}} z(x-u)F(du) \tag{8}$$

satisfies $z(x) \leq z(0)$ for every $x \in \mathbb{R}$, then z is a constant.

Proof. As we have

$$z(0) = \int_0^\infty z(-u)F^{\ell *}(du),$$

for $\ell = 1, 2, \ldots$, and $z(-u) \leq z(0)$, then $z(-y) = z(0)$ holds for $F^{\ell *}$-a.a. y and, for all $\ell = 1, 2, \ldots$. Since F is non-arithmetic, for any $\epsilon > 0$ there is x_0 such that for any $x > x_0$ the interval $(x, x+\epsilon)$ contains a

point of increase of $F^{\ell *}$ for some $\ell \geq 1$ (by Lemma 4.4). By the uniform continuity of z, we get $z(-y) \to z(0)$ as $y \to \infty$. Due to (6), for any $x \geq 0$, we have $F^{\ell *}(x) \to 0$ as $\ell \to \infty$. Hence, we get

$$z(x) - z(0) = \int_0^\infty (z(x-u) - z(0)) F^{\ell *}(du) \to 0$$

as $\ell \to \infty$. □

Let f be a real valued function defined on $[0, \infty)$. Then f is called **directly integrable** if the following conditions are met. If we set $m_k(h) = \inf_{(k-1)h \leq x \leq kh} f(x)$, plus $M_k(h) = \sup_{(k-1)h \leq x \leq kh} f(x)$ for $h > 0$, and $s(h) = h \sum_k m_k(h)$, $S(h) = h \sum_k M_k(h)$, then $s(h), S(h)$ are absolutely convergent and we have $\lim_{h \to 0}(S(h) - s(h)) = 0$.

Theorem 4.6 (Renewal theorem)**.** *Assume that F is a non-arithmetic distribution function on $(0, \infty)$ and $y : [0, \infty) \to \mathbb{R}$ is directly integrable. Then for any solution z of the renewal equation* (7)*, we have that*

$$\lim_{x \to \infty} z(x) = \frac{1}{\mu} \int_0^\infty y(u) du$$

holds with $\mu = \int_0^\infty x F(dx)$.

Proof. The definition of K gives $\int_0^\infty (K(x) - K(x-y)) F^{\ell *}(dy) \leq \ell$. For any $\alpha > 0$, there is $\ell \geq 1$ such that $F^{\ell *}(\alpha) < 1$. Then for $x > \alpha$ we have

$$\ell \geq \int_\alpha^\infty (K(x) - K(x-y) F^{\ell *}(dy) \geq (1 - F^{\ell *}(\alpha))(K(x) - K(x-\alpha)).$$

By this inequality, for each fixed interval I, we get $\sup_{t \geq 0} K(I+t) < \infty$. By Helly's selection theorem there exists a sequence $\{t_k\}$ and a measure L on $\mathbb{R}$ such that $\lim_{k \to \infty} t_k = \infty$ and $K_{t_k}((a,b]) = K((a,b] + t_k) \to L((a,b])$ at the continuity points of L. Let w be a continuous function which vanishes on $[0,a]^c$. If we set $z_w(x) = \int_{-\infty}^\infty w(x-s) K(ds)$, we get

$$z_w(t_k + x) = \int_{-\infty}^\infty w(x-s) K_{t_k}(ds) \to \int_{-\infty}^\infty w(x-s) L(ds) = \zeta(x).$$

Note that the function z_w satisfies the renewal equation

$$z_w(t_k + x) = w(t_k + x) + \int_0^\infty z_w(t_k + x - s)F(ds).$$

The left hand side of the above equality tends to $\zeta(x)$ as $k \to \infty$, while the right hand side tends to $\int_0^\infty \zeta(x-s)F(ds)$ by bounded convergence. Thus ζ satisfies (8). Since ζ is bounded and uniformly continuous, by (8), we get $\zeta(x) = \zeta(0)$. So L is a translation invariant measure and therefore is a multiple of Lebesgue measure. Hence there exists $\beta > 0$ such that

$$K(t_k) - K(t_k - h) \to \beta h.$$

Further, note that

$$1 = \int_0^\infty K((t_k - u, t_k])F(du) \to \beta\mu$$

implies $\beta = \frac{1}{\mu}$. Owing to this fact, we obtain $K(t) - K(t-h) \to \frac{h}{\mu}$ as $t \to \infty$. Since y is directly integrable, denoting by M_n the supremum and by m_n the infimum of $y(u)$ on $((n-1)h, nh]$, we have

$$\sum_n m_n K([t_k - nh, t_k - (n-1)h)) \le z(t_k) = \int_{-\infty}^0 y(-u)K_{t_k}(du)$$
$$\le \sum_n M_n K([t_k - nh, t_k - (n-1)h)).$$

By bounded convergence, the left and right tend to $\frac{h}{\mu}\sum_{n=1}^\infty m_n$ and $\frac{h}{\mu}\sum_{n=1}^\infty M_n$, respectively, as $k \to \infty$. As $h \to 0$, each side reaches $\frac{1}{\mu}\int_0^\infty y(u)du$. □

Lemma 4.7. *Let $f(x)$ be a nonnegative and non-increasing function satisfying $\int_0^\infty f(x)e^{Rx}dx < \infty$ for $R > 0$. Then $f(x)e^{Rx}$ is directly integrable on $[0, \infty)$.*

Proof. For $s(h) = h\{f(h) + f(2h)e^{Rh} + \cdots\}$ and $S(h) = h\{f(0)e^{Rh} + f(h)e^{2Rh} + \cdots\}$ we have $s(h) \le \int_0^\infty f(x)e^{Rx}dx < \infty$ and $S(h) - s(h) = hf(0)e^{Rh} + (e^{2Rh} - 1)s(h)$, hence also $\lim_{h\to 0}(S(h) - s(h)) = 0$. □

Theorem 4.8. *If a distribution function $F(x)$ on $(0,\infty)$ satisfies the Cramér-Lundberg condition and*

$$\mu^* = \frac{\lambda}{c}\int_0^\infty e^{r^*x}x(1-F(x))dx < \infty$$

holds, then

$$\lim_{u\to\infty} e^{r^*u}\mathcal{E}(u) = \frac{1}{\mu^* r^*}(1-\frac{\lambda\mu}{c}).$$

Proof. By (4), we get

$$\varphi(u) = \varphi(0) + \frac{\lambda}{c}\int_0^u \varphi(u-z)(1-F(z))dz,$$

or what is the same

$$\begin{aligned} 1-\varphi(u) &= 1-\varphi(0)-\frac{\lambda}{c}\int_0^u (1-F(z))dz \\ &\quad +\frac{\lambda}{c}\int_0^u (1-\varphi(u-z))(1-F(z))dz. \end{aligned}$$

From this, it follows

$$\begin{aligned} e^{r^*u}\mathcal{E}(u) &= \frac{\lambda}{c}(\mu - \int_0^u (1-F(z))dz)e^{r^*u} \\ &\quad + \int_0^u \mathcal{E}(u-z)e^{r^*(u-z)}\frac{\lambda}{c}e^{r^*z}(1-F(z))dz. \end{aligned}$$

Therefore the above is a renewal equation on $e^{r^*u}\mathcal{E}(u)$. We will now apply the renewal theorem, Theorem 4.6, to this situation. For this, note that the distribution $\frac{\lambda}{c}e^{r^*z}(1-F(z))dz$ is non-arithmetic. Next, by the Cramér-Lundberg condition we get $\int_0^\infty \frac{\lambda}{c}(1-F(z))e^{r^*z}dz = 1$, and so we have

$$\begin{aligned} &\frac{\lambda}{c}\int_0^\infty (\mu - \Big(\int_0^u (1-F(z))dz\Big)e^{r^*u}du \\ &\qquad = \frac{\lambda}{c}\int_0^\infty \Big(\int_u^\infty (1-F(z))dz\Big)e^{r^*u}du \end{aligned}$$

$$
\begin{aligned}
&= \frac{\lambda}{c}\int_0^\infty (1-F(z))\Big(\int_0^z e^{r^*u}du\Big)dz \\
&= \frac{\lambda}{c}\int_0^\infty (1-F(z))\frac{e^{r^*z}-1}{r^*}dz \\
&= \frac{1}{r^*}(1-\frac{\lambda\mu}{c}).
\end{aligned}
$$

Therefore we can apply the renewal theorem in order to get

$$\lim_{u\to\infty} e^{r^*u}\mathcal{E}(u) = \frac{1}{\mu^* r^*}\Big(1-\frac{\lambda\mu}{c}\Big).$$

□

Example 4.9. Let $F(x) = e^{-x/\mu}$. Then the Cramér-Lundberg exponent is $r^* = \frac{\rho}{\mu(1+\rho)}$ and we have $\mu^* = \frac{c}{\lambda}$. Hence $\frac{1}{\mu^* r^*}(1-\frac{\lambda\mu}{c}) = \frac{1}{1+\rho}$. By Example 3.4, we must have $e^{r^*u}\mathcal{E}(x) = \frac{1}{1+\rho}$.

4.2 Asymptotics of ruin probability (heavy tailed claims)

In this section we discuss the case when the tail of each claim does not satisfy the Cramér-Lundberg condition. In recent years, big disasters had occurred. These phenomena imply the necessity of discussing the heavy tailed claims which cannot be treated by the Cramér-Lundberg approximation. A typical example of such heavy tailed claim is a regularly varying tail. So, at first, we explain necessary basic facts on distributions with regularly varying tails. Then we talk about subexponential distributions where the distributions with regularly varying tails are contained (see Lemma 4.25(i)). For these subexponential distributions the asymptotics of ruin probability can be easily derived.

Regular variation

A measurable function $\ell : [0, \infty) \to (0, \infty)$ is called **slowly varying at** ∞ if for any $y > 0$ we have

$$\lim_{x\to\infty} \frac{\ell(xy)}{\ell(x)} = 1.$$

We will use $h(x) = \log \ell(e^x)$ in what follows.

Theorem 4.10 (Uniform convergence theorem)**.** *If a function ℓ is slowly varying at infinity, then for any $A > 0$ we have*

$$\lim_{x\to\infty} \sup_{0\le y\le A} \left| \frac{\ell(xy)}{\ell(x)} - 1 \right| = 0.$$

Proof. We will verify $\sup_{0\le u\le A} |h(x + u) - h(x)| = 0$ for a given $A > 0$. Choose ϵ so that $0 < \epsilon < A$ and define

$$\begin{aligned} I_x &= [x, x + 2A], \\ E_x &= \{t \in I_x : |h(t) - h(x)| \ge \frac{\epsilon}{2}\}, \\ E_x^* &= \{t \in [0, 2A] : |h(x + t) - h(x)| \ge \frac{\epsilon}{2}\}. \end{aligned}$$

The sets E_x, E_x^* are Borel measurable and $|E_x| = |E_x^*|$ (here $|\cdot|$ denotes Lebesgue measure). By the definition of slow variation, for any $y \ge 0$ we have $\lim_{x\to\infty} 1_{E_x^*}(y) = 0$. Hence, the dominated convergence theorem delivers $|E_x^*| \to 0$ as $x \to \infty$. This yields the existence of x_0 for which $x \ge x_0$ implies $|E_x| < \frac{\epsilon}{2}$. For $c \in [0, A]$ we have $I_x \cap I_{x+c} = [x+c, x+2A]$, while $|I_x \cap I_{x+c}| = 2A - c \ge A$ for $c \le A$. In this way, for $x \ge x_0$, we obtain

$$|E_x \cup E_{x+c}| < |E_x| + |E_{x+c}| = |E_x^*| + |E_{x+c}^*| < \epsilon < A.$$

Hence, given $c \in [0, A]$ and $x \ge x_0$ we get

$$|(I_x \cap I_{x+c}) \backslash (E_x \cup E_{x+c})| > 0.$$

And we conclude $J_x = (I_x \cap I_{x+c})\backslash(E_x \cup E_{x+c}) \neq \emptyset$. Since all $t \in J_x$ satisfies

$$\begin{aligned} |h(t) - h(x)| &< \epsilon/2 \\ |h(t) - h(x+c)| &< \epsilon/2, \end{aligned}$$

at the end we get $|h(x+c) - h(x)| < \epsilon$. □

Lemma 4.11. *If $\ell : [A, \infty) \to (0, \infty)$ satisfies*

$$\lim_{x\to\infty} \ell(xy)/\ell(x) = 1$$

for any $y > 0$, then ℓ and h are bounded in any bounded interval sufficiently far away from the origin.

Proof. By the uniform convergence theorem (Theorem 4.10) there is x_0 such that for all $x \geq x_0$ we have

$$\sup_{0\leq u\leq 1} |h(x+u) - h(x)| < 1.$$

Hence we get

$$\begin{aligned} |h(x)| &\leq 1 + |h(x_0)| \quad on \quad [x_0, x_0+1], \\ &\vdots \\ |h(x)| &\leq n + |h(x_0)| \quad on \quad [x_0, x_0+n]. \end{aligned}$$

□

Remark 4.12. Due to this lemma, a slowly varying function is integrable on finite intervals sufficiently far away from the origin.

Theorem 4.13 (Representation theorem)**.** *A function ℓ is slowly varying at infinity if and only if it can be represented as*

$$\ell(x) = c(x) \exp\{\int_a^x \frac{e(\log u)}{u} du\} \quad (x \geq a > 0),$$

where c and e are measurable functions subject to $\lim_{x\to\infty} c(x) > 0$ and $\lim_{x\to\infty} e(x) = 0$, respectively.

Proof. Choose x_0 large enough so that ℓ is integrable in any finite interval in $[x_0, \infty)$. For $x \geq x_0$, h is written as

$$h(x) = \int_x^{x+1} \{h(x) - h(t)\}dt + \int_{x_0}^{x} \{h(t+1) - h(t)\}dt + \int_{x_0}^{x_0+1} h(t)dt.$$

Set $e(x) = h(x+1) - h(x)$. Then we have $\lim_{x\to\infty} e(x) = 0$. From the uniform convergence theorem we get

$$\begin{aligned} \left| \int_x^{x+1} \{h(x) - h(t)\}dt \right| &= \left| \int_0^1 \{h(x) - h(x+u)\}du \right| \\ &\leq \int_0^1 \sup_{0\leq u\leq 1} |h(x) - h(x+u)|du \to 0 \end{aligned}$$

as $(x \to \infty)$. Set $c = \int_{x_0}^{x_0+1} h(t)dt$ and $d(x) = c + \int_x^{x+1} \{h(x) - h(t)\}dt$. Then we have

$$h(x) = d(x) + \int_{x_0}^{x} e(t)dt,$$

and so also

$$\ell(x) = e^{h(\log x)} = e^{d(\log x)} \exp\{\int_{x_0}^{x} \frac{e(\log v)}{v} dv\}.$$

□

Lemma 4.14 (Hamel equation)**.** *If a measurable function f satisfies $f(x+y) = f(x) + f(y)$ for all $x, y \in \mathbb{R}$, then there is $c \in \mathbb{R}$ such that f is represented as $f(x) = cx$.*

Proof. We omit the proof. We refer the interested reader to [3, 5]. □

Theorem 4.15. *Let $f : [0, \infty) \to (0, \infty)$ be measurable. Suppose that the limit*

$$g(y) = \lim_{x\to\infty} \frac{f(xy)}{f(x)} > 0$$

exists for every $y > 0$. Then for some $\rho \in \mathbb{R}$ we have $g(y) = y^\rho$ for $y > 0$. Furthermore, there exists a function ℓ slowly varying at infinity such that $f(x) = x^\rho \ell(x)$.

Proof. For y, z positive, we have

$$\frac{f(xyz)}{f(x)} = \frac{f(xyz)}{f(xy)}\frac{f(xy)}{f(x)}.$$

Therefore by taking limits we obtain $g(zy) = g(z)g(y)$. Set $h(x) = \log g(e^x)$. Then for $x, y \in \mathbb{R}$ we get

$$h(x+y) = \log g(e^x e^y) = \log g(e^x) + \log g(e^y) = h(x) + h(y).$$

By Lemma 4.14, there is $\rho \in \mathbb{R}$ such that $h(x) = \rho x$. Hence, one has $g(y) = e^{h(\log y)} = y^\rho$.

For the second part, it is enough to define $\ell(x) = f(x)/x^\rho$. In fact, as we have

$$\begin{aligned}\frac{\ell(xy)}{\ell(x)} &= \frac{f(xy)}{(xy)^\rho}\frac{x^\rho}{f(x)} \\ &= y^{-\rho}\frac{f(xy)}{f(x)} \to 1 \quad (x \to \infty),\end{aligned}$$

this function ℓ is slowly varying at infinity. □

Remark 4.16. A positive function near infinity which satisfies the second part of the above theorem is called **regularly varying at infinity with exponent** ρ.

Lemma 4.17 (Continuity theorem)**.** *For $n \in \mathbb{N}$, let F_n be a probability measure on $[0, \infty)$ and ω_n be its Laplace transform.*
(1) *If $\lim_{n\to\infty} \omega_n(s) = \omega(s)$ for $s > 0$, then ω is the Laplace transform of a measure F with total mass not greater than 1 and $\lim_{n\to\infty} F_n(x) = F(x)$ for any continuity point x of F.*
(2) *If $\lim_{n\to\infty} F_n(x) = F(x)$ holds for any continuity point x of a measure F with total mass not greater than 1, then for $s > 0$ we have $\lim_{n\to\infty} \omega_n(s) = \omega(s)$.*

Proof. See [3, page 431]. □

Lemma 4.18 (Extended continuity theorem)**.** *For $n \in \mathbb{N}$, let U_n be a measure on $[0,\infty)$ and ω_n be its Laplace transform.*
(1) *If for some $a > 0$ we get $\lim_{n\to\infty} \omega_n(s) = \omega(s)$ for any $s > a$, then ω is the Laplace transform of a measure U and it satisfies $\lim_{n\to\infty} U_n(I) = U(I)$ for any continuity interval I.*
(2) *If for any continuity interval I of U conditions $\lim_{n\to\infty} U_n(I) = U(I)$ and $\{\omega_n(a)\}_{n\geq 1}$ is bounded are satisfied, then we have $\lim_{n\to\infty} \omega_n(s) = \omega(s)$ for $s > a$.*

Proof. (1) Choose $s_0 > a$. Note that $U_n^0(dx) = \omega_n(s_0)^{-1} e^{-s_0 x} U_n(dx)$ is a probability measure and its Laplace transform is $\omega_n(s + s_0)/\omega_n(s_0)$. By the continuity theorem, there exists a measure U^0 with total mass not greater than 1 such that U_n^0 converges to U^0. Hence $U_n(dx) = \omega(s_0) e^{s_0 x} U_n^0(dx)$ converges to $\omega(s_0) e^{s_0 x} U^0(dx)$.
(2) Let $t > 0$ be a continuity point of U. Then one obtains

$$\lim_{n\to\infty} \int_0^t e^{-sx} U_n(dx) = \int_0^t e^{-sx} U(dx).$$

Furthermore, we have

$$\int_t^\infty e^{-sx} U_n(dx) \leq e^{-(s-a)t} \int_0^\infty e^{-ax} U_n(dx) < A e^{-(s-a)t} \to 0 \ (s > a)$$

as $t \to \infty$, where $A = \sup_n \omega_n(a)$. For U, a similar property applies. Hence, one finally obtains $\lim_{n\to\infty} \omega_n(s) = \omega(s)$. □

Theorem 4.19 (Tauberian theorem)**.** *For a non-decreasing right continuous function $U : [0,\infty) \to (0,\infty)$, the following properties*
(1) $U(x) \sim c x^\rho \ell(x)/\Gamma(1+\rho) \quad (x \to \infty)$,
(2) $\int_{[0,\infty)} e^{-sx} dU(x) \sim c s^{-\rho} \ell(1/s) \quad (s \downarrow 0)$,
are equivalent; here ℓ is slowly varying function at infinity and $c, \rho \geq 0$.

Proof. Let ω be the Laplace transform of U.

We first proof that (2) implies (1). By (2), one has $\frac{\omega(st)}{\omega(s)} \to t^{-\rho}$ as $s \to 0$. Let $U_{1/s}(x) = U(x/s)$. Then $\frac{\omega(st)}{\omega(s)}$ and $t^{-\rho}$ are the Laplace

transforms of the measures $\frac{U_{1/s}(dx)}{\omega(s)}$ and $\frac{x^{\rho-1}}{\Gamma(\rho)}dx$, respectively. Applying the extended continuity theorem, one has that the measure $\frac{U_{1/s}(dx)}{\omega(s)}$ converges to $\frac{x^{\rho}}{\Gamma(\rho)}dx$. By the assumption, we obtain $U(x) \sim cx^{\rho}\ell(x)/\Gamma(1+\rho)$.

Reciprocally, by assumption we have $\lim_{x\to\infty}\frac{U(tx)}{U(x)} = t^{\rho}$. The Laplace transform of $\frac{U(tx)}{U(x)}$ is $\frac{\omega(s/x)}{U(x)}$ and $\int_0^{\infty} e^{-st}d(t^{\rho})$ equals $s^{-\rho}\Gamma(\rho+1)$. Dividing the interval of integration, we have $\omega(1/x) \leq U(x) + \sum_{n=1}^{\infty} e^{-2^{n-1}}U(2^n x)$. We can choose x_0 so that $U(2x) < 2^{\rho+1}U(x)$ for $x > x_0$. Thus we obtain $U(2^n x) < 2^{n(\rho+1)}U(x)$. Using this gives

$$\frac{\omega(1/x)}{U(x)} \leq 1 + \sum_{n=1}^{\infty} e^{-2^{n-1}}2^{n(\rho+1)} < \infty.$$

Hence we get $\lim_{x\to\infty}\frac{\omega(s/x)}{U(x)} = s^{-\rho}\Gamma(\rho+1)$ by the extended continuity theorem. By assumption (1), we finally have $\omega(s) \sim cs^{-\rho}\ell(1/s)$. □

Subexponentiality

Define $\overline{F} = 1 - F$ for a distribution function F on $[0,\infty)$.

A distribution function F on $[0,\infty)$ is called **subexponential** if it satisfies

$$\lim_{x\to\infty}\overline{F^{2*}}(x)/\overline{F}(x) = 2.$$

We denote the class of subexponential distribution functions by $\mathcal{S}$.

Lemma 4.20. *For $F \in \mathcal{S}$ the following two properties are satisfied.*
(1) *For any $A > 0$, we have*

$$\lim_{x\to\infty}\sup_{0\leq y\leq A}\left|\frac{\overline{F}(x-y)}{\overline{F}(x)} - 1\right| = 0,$$

(2) *For any $\epsilon > 0$, we have*

$$\lim_{x\to\infty} e^{\epsilon x}\overline{F}(x) = \infty.$$

Proof. For the first part, we fix $t \in (0, x)$ in order to have

$$\begin{aligned}
\overline{F^{2*}}(x) &= 1 - F^{2*}(x) = 1 - \int_0^x F(x-y)dF(y) \\
&= 1 - F(x) + \int_0^x \overline{F}(x-y)dF(y) \\
&= \overline{F}(x) + \Big(\int_0^t + \int_t^x\Big)\overline{F}(x-y)dF(y) \\
&\geq \overline{F}(x) + \overline{F}(x)F(t) + (F(x) - F(t))\overline{F}(x-t).
\end{aligned}$$

Divide both sides by $\overline{F}(x)$. Since

$$\frac{\overline{F^{2*}}(x)}{\overline{F}(x)} \geq 1 + F(t) + \frac{(F(x) - F(t))\overline{F}(x-t)}{\overline{F}(x)}$$

is satisfied, we have

$$\Big(\frac{\overline{F^{2*}}(x)}{\overline{F}(x)} - 1 - F(t)\Big) \Big/ (F(x) - F(t)) \geq \frac{\overline{F}(x-t)}{\overline{F}(x)}.$$

The left hand side of the above equality converges to 1 as $x \to \infty$. One has then $\limsup_{x\to\infty} \frac{\overline{F}(x-t)}{\overline{F}(x)} \leq 1$. Obviously we have $\frac{\overline{F}(x-t)}{\overline{F}(x)} \geq 1$ and therefore also

$$\lim_{x\to\infty} \frac{\overline{F}(x-t)}{\overline{F}(x)} = 1.$$

Since

$$1 \leq \frac{\overline{F}(x-t)}{\overline{F}(x)} \leq \frac{\overline{F}(x-A)}{\overline{F}(x)}$$

holds for $\leq t \leq A$, the result follows.

Now we prove the second part. By (1), the distribution $\overline{F}(\log u)$ is slowly varying at infinity. By the representation theorem (Theorem 4.13), for any $\epsilon > 0$ we have

$$u^{\epsilon}\overline{F}(\log u) = e^{\int_1^u \frac{\epsilon}{v}dv}\overline{F}(\log u) \to \infty$$

as $u \to \infty$. □

Lemma 4.21. *Let F be a distribution function on $[0,\infty)$. If $F(x)>0$ for any $x>0$, then we have*

$$\liminf_{x\to\infty}\frac{\overline{F^{n*}}(x)}{\overline{F}(x)}\geq n.$$

Proof. Note that $F^{n*}(x)\leq F^n(x)$ holds by (6). Therefore we have

$$\frac{\overline{F^{n*}(x)}}{\overline{F}(x)}\geq\frac{1-F^n(x)}{1-F(x)}=\sum_{k=0}^{n-1}F^k(x)\to n\ (x\to\infty).$$

□

Lemma 4.22. *If $F\in\mathcal{S}$, then for every $n\geq 1$ we have*

$$\lim_{x\to\infty}\overline{F^{n*}}(x)/\overline{F}(x)=n.$$

Proof. We show the result for $n=3$. First note the equalities

$$\begin{aligned}1-F^{3*}(x) &= 1-\int_0^x F^{2*}(x-y)dF(y)\\ &= 1-F(x)+\int_0^x\{1-F^{2*}(x-y)\}dF(y),\end{aligned}$$

$$\frac{1-F^{3*}(x)}{1-F(x)}=1+\int_0^x\frac{1-F^{2*}(x-y)}{1-F(x-y)}\frac{1-F(x-y)}{1-F(x)}dF(y).$$

By subexponentiality of F and Lemma 4.20 part (1) the integrand on the right hand side of the latter equality tends to 2 for any $y\geq 0$ as $x\to\infty$. We have for $t\in(0,x)$ then

$$\begin{aligned}&\Big(\int_0^t+\int_t^x\Big)\frac{1-F^{2*}(x-y)}{1-F(x-y)}\frac{1-F(x-y)}{1-F(x)}dF(y) \qquad (9)\\ \leq&\sup_{x-t\leq y\leq x}\frac{1-F^{2*}(y)}{1-F(y)}\frac{1-F(x-t)}{1-F(x)}F(t)+K\int_t^x\frac{1-F(x-y)}{1-F(x)}dF(y),\end{aligned}$$

here $K=\sup_{y\geq 0}\frac{1-F^{2*}(y)}{1-F(y)}<\infty$. Our goal is to find the limit of (9) in order to get an upper bound. Since we have $\lim_{x\to\infty}\frac{\overline{F}^{2*}(x)}{\overline{F}(x)}=2$ and

$\dfrac{\overline{F}^{2*}(x)}{\overline{F}(x)} \leq 2$, we obtain $\sup\limits_{x-t\leq y\leq x} \dfrac{1-F^{2*}(y)}{1-F(y)} \to 2$ as $x \to \infty$. Moreover, we have

$$\begin{aligned} F(t) &\leq \int_0^t \frac{1-F(x-y)}{1-F(x)} dF(y) \\ &\leq \sup_{0\leq y\leq t} \frac{1-F(x-y)}{1-F(x)} F(t) \to F(t) \end{aligned}$$

and

$$\int_0^x \frac{1-F(x-y)}{1-F(x)} dF(y) = \frac{F(x)-1+1-F^{2*}(x)}{1-F(x)} \to -1+2 = 1.$$

Hence, we get

$$\int_t^x \frac{1-F(x-y)}{1-F(x)} dF(y) \to 1-F(t)$$

as $x \to \infty$. By letting $x \to \infty$, the right hand side of (9) converges to $2F(t) + K(1-F(t))$. Next, by letting $t \to \infty$, this term converges to 2. With Lemma 4.21 at hand we conclude

$$\frac{1-F^{3*}(x)}{1-F(x)} \to 3 \quad \text{as} \quad x \to \infty.$$

□

Lemma 4.23. *If a distribution function F on $[0,\infty)$ is subexponential, then for any $\epsilon > 0$ there is $K > 0$ such that for all $x \geq 0$ we have*

$$\overline{F^{n*}}(x)/\overline{F}(x) \leq K(1+\epsilon)^n, \ n = 2, 3, \ldots.$$

Proof. Define $\alpha_n = \sup_x \overline{F^{n*}}(x)/\overline{F}(x)$. Then we have

$$\frac{\overline{F^{(n+1)*}}(x)}{\overline{F}(x)} = 1 + \int_0^x \frac{1-F^{n*}(x-y)}{1-F(x)} F(dy)$$

$$
\begin{aligned}
&\leq 1+\sup_{x\leq M}\int_0^x \frac{1}{1-F(x)}F(dy)+\\
&\quad +\sup_{x\geq M}\int_0^x \frac{1-F^{n*}(x-y)}{1-F(x-y)}\frac{1-F(x-y)}{1-F(x)}F(dy)\\
&\leq 1+\frac{1}{1-F(M)}+\\
&\quad +\sup_x \frac{1-F^{n*}(x)}{1-F(x)}\times\sup_{x\geq M}\int_0^x \frac{1-F(x-y)}{1-F(x)}F(dy)\\
&\leq 1+\frac{1}{1-F(M)}+\alpha_n(1+\epsilon).
\end{aligned}
$$

Where we have used that $|\sup_{x\geq M}\int_0^x \frac{1-F(x-y)}{1-F(x)}F(dy)-1|<\epsilon$ works for sufficiently big M. Therefore we conclude

$$
\alpha_{n+1}\leq\Big(1+\frac{1}{1-F(M)}\Big)\frac{1}{\epsilon}(1+\epsilon)^{n+1}.
$$

□

Lemma 4.24. *Let F be a subexponential distribution function on $[0,\infty)$. Assume that the power series $\sum_{n=0}^\infty a_n s^n$ is analytic at $s=1$, where $a_n\geq 0$, $n=0,1,2,\dots$ and satisfies $\sum_{n=0}^\infty a_n=1$. If we set $G(x)=\sum_{n=0}^\infty a_n F^{n*}(x)$, then we get*

$$
\lim_{x\to\infty}\frac{\overline{G}(x)}{\overline{F}(x)}=\sum_{n=1}^\infty na_n.
$$

Proof. The result is obvious from Lemmas 4.22, 4.23 and the dominated convergence theorem. □

Lemma 4.25. *Let F be a distribution function on $(0,\infty)$.*
(1) *If $1-F(x)\sim x^{-\rho}\ell(x)$ for a function ℓ slowly varying at infinity, then $F\in\mathcal{S}$.*
(2) *If F satisfies the Cramér-Lundberg condition, then $F\notin\mathcal{S}$.*

Proof. (1) Assume that the two positive random variables X_1 and X_2 are independent with common distribution function F. For any $\epsilon > 0$ we have

$$\begin{aligned} P(X_1 + X_2 \geq z) &\leq P(X_1 \geq (1-\epsilon)z) + P(X_2 \geq (1-\epsilon)z) \\ &\quad + P(X_1 > \epsilon z, X_2 > \epsilon z) \\ &\sim 2(1-\epsilon)^{-\rho} z^{-\rho} \ell(z) + \{\epsilon^{-\rho} z^{-\rho} \ell(z)\}^2. \end{aligned}$$

Hence, one obtains $\limsup_{z\to\infty} \frac{P(X_1+X_2\geq z)}{P(X_1\geq z)} \leq 2(1-\epsilon)^{-\rho}$. Since ϵ is arbitrary, the inequality

$$\limsup_{z\to\infty} \frac{P(X_1 + X_2 \geq z)}{P(X_1 \geq z)} \leq 2$$

follows. On the other hand, using $P(X_1 + X_2 \geq z) \geq P(X_1 \geq z) + P(X_2 \geq z) - P(X_1 \geq z, X_2 \geq z)$ we have

$$\liminf_{z\to\infty} \frac{P(X_1 + X_2 \geq z)}{P(X_1 \geq z)} \geq 2.$$

(2) If $F \in \mathcal{S}$, then by Lemma 4.20, part (2), for any $\epsilon > 0$ we have $e^{\epsilon x}\overline{F}(x) \to \infty$ as $x \to \infty$. So, for every $R > 0$, we get $\int_0^\infty e^{Rx}(1 - F(x))dx = \infty$, which leads to a contradiction. □

Finally we have the following characterization of the ruin probability. Recall that we have assumed $\frac{c}{\lambda\mu} > 1$, i.e., $\rho > 0$.

Theorem 4.26. *Let F be a distribution function on $(0, \infty)$ with mean $\mu < \infty$. Then $H(x) = \int_0^x \frac{1-F(y)}{\mu} dy \in \mathcal{S}$ implies*

$$\mathcal{E}(u) \sim \frac{1}{\rho}\overline{H}(u) \text{ as } u \to \infty.$$

Proof. By Theorem 3.3 we have

$$\mathcal{E}(u) = \frac{\rho}{1+\rho} \sum_{n=0}^{\infty} \frac{\overline{H^{n*}}(u)}{(1+\rho)^n}.$$

By Lemma 4.24 we conclude

$$\begin{aligned}\mathcal{E}(u) &\sim \frac{\rho}{1+\rho}\sum_{n=0}^{\infty}\frac{n}{(1+\rho)^n}\overline{H}(u)\\ &= \frac{1}{\rho}\overline{H}(u)\end{aligned}$$

as $u \to \infty$. □

In contrast with the light tail case, using the previous theorem, we easily obtain the behavior as $u \to \infty$ of $\mathcal{E}(u)$ if $\overline{F}$ is regularly varying at infinity.

Acknowledgments: During the preparation of this note, Professor Kohatsu-Higa and the editor gave the author valuable advices. The author is indebted to them.

References

[1] Chistyakov, V.P.; *A theorem on sums of independent positive random variables and its applications to branching processes.* Theory Probability Appl. 9 (1964) 640-648.

[2] Embrechts, P., Goldie, C. M., Veraverbeke, N.; *Subexponentiality and infinite divisibility,* Z. Wahrscheinlichkeitstheorie verw. Gebiette 49 (1979) 335-347.

[3] Feller, W.; *An Introduction to Probability Theory and its Applications, vol.2. 2nd edition.*, John Wiley & Sons (1971).

[4] Kendall, D. G.; *Some problems in the theory of dams*, J. Roy. Statist. Soc. Ser. B. 19 (1957) 207-212.

[5] Kuczma, M.; *An Introduction to the Theory of Functional Equations and Inequalities*, Birkhauser, (2009), 2nd edition.

[6] Mikosch, T.; *Non-Life Insurance Mathematics*, Springer (2006).

[7] Sato, K.; *Lévy processes and infinitely divisible distributions*, Cambridge University Press, Cambridge (1999).

Resumen

En este artículo describimos los conceptos básicos relacionados a seguros que no sean de vida y luego explicamos procesos de riesgo. En particular, tratamos al detalle el comportamiento asintótico de la probabilidad de que un producto sea declarado en ruina. Como es suponible, el comportamiento en el horizonte depende de la cola de la distribución de las primas.

Palabras clave: Procesos estocásticos, matemáticas actuariales, seguros no de vida, probabilidad de ruina, aproximación de Cramér-Lundberg.

Makoto Yamazato
Department of Mathematical Sciences,
University of the Ryukyus
Senbaru, Nishihara-cho, Okinawa 903-0213,
Japan
makotoyamazato@yahoo.co.jp

Información para los autores

La revista PRO MATHEMATICA publica artículos originales de investigación, así como trabajos de divulgación, en todas las áreas de matemáticas.

El manejo de la revista está a cargo de un Director, un Consejo Directivo y un Consejo Editorial; el Director por definición pertenece al Consejo Directivo, y los miembros del Consejo Directivo, a su vez, son parte del Consejo Editorial.

Procedimiento editorial. Los autores interesados en publicar en PRO MATHEMATICA deberán someter sus manuscritos originales en español o inglés vía correo electrónico en formato .PDF a un miembro del Consejo Editorial, con copia a **promathematica@pucp.edu.pe** e indicando como asunto **artículo para ProMathematica**.
El editor dentro de las siguientes dos semanas revisará someramente el material y presentará un breve informe al Consejo Directivo. En base a este informe el Consejo procederá a rechazar el artículo o a someterlo a un arbitraje por pares, lo cual será informado al autor.
En caso el Consejo dictamine que el contenido del artículo lo amerita, se procederá a enviarlo a revisión por un experto en el área elegido en coordinación con el editor que lo elevó. La comunicación con el revisor será a través del editor receptor del artículo. Dentro de los siguientes tres meses el experto enviará un informe al Consejo Directivo indicando expresamente si el artículo es rechazado, presenta observaciones de fondo o es aceptado con observaciones menores (o sin ellas). El Consejo Editorial enviará al autor una copia del informe para que actúe acorde. El autor en esta etapa procederá a levantar las objeciones y a realizar las modificaciones que juzgue necesarias. El artículo revisado será reenviado al revisor para que indique su conformidad o no con la nueva versión.

Preparación para publicación. Una vez aceptado el artículo para publicación, los autores enviarán una copia en LaTeX para que sea editado de acuerdo con el estilo de la revista. Para evitar imponer en los autores una excesiva carga, la revista se encargará del diagramado y las revisión de estilo, excepto en casos cuando los autores prefieran realizarlo por su cuenta. Terminado el proceso, se pedirá a los autores que lean cuidadosamente las pruebas de imprenta e indiquen posibles observaciones al trabajo de edición.

Compromiso de integridad. El envío de un manuscrito conlleva un compromiso por parte de los autores de que (1) no ha sido publicado anteriormente, (2) no ha sido enviado a otra revista, (3) su publicación es aprobado por el autor y todos los coautores y (4) tiene autorización para utilizar todo material protegido por derechos de autor que contenga.

Toda correspondencia deberá esta dirigida a

PRO MATHEMATICA
Pontificia Universidad Católica del Perú
Departamento de Ciencias
Apartado 1761
Lima 100, Perú

correo electrónico: promathematica@pucp.edu.pe

Information for authors

PRO MATHEMATICA publishes research and survey papers in all math areas.

The Journal is managed by a Director, a Consulting Board, and an Editorial Board. The Director belongs to the Consulting Board, whose members in turn are part of the Editorial Board.

Editorial procedure. Authors interested in publishing in PRO MATHEMATICA should submit their original work in .PDF format, either in English or Spanish, via email addressed to a member of the Editorial Board, copied to **promathematica@pucp.edu.pe**, with subject indicated as **paper for ProMathematica**.
The Editor within the next two weeks will present a brief report to the Editorial Board. Based on this report, the Board will decide upon rejecting the article or submitting it to a thorough peer review. Either way, the author will be notified.
In case the consulting Board decides that the content of the material merits it, the article will be send to an expert, chosen in coordination with the reporting editor. All communication between the reviser and the author is carried through the Editor. Within the next three months, the reviser would report the Consulting Board, indicating if the article should be rejected or if it should be accepted with minor or major observations. The Board will send the author a copy of the report so that he can act accordingly. At this stage the author must revise his work and make the necessary modifications. Once resubmitted, the paper will be sent to the referee so that he can give his conformity.

Preparation for publication. Once accepted for publication, the authors would send a LaTeX source file so that it can be edited according to the journal's style. In order not to impose the authors with unnecessary burden, our staff would take care of the diagramming and style review, except in case the authors prefer to do it themselves. Once finished the process, the authors would be asked to carefully proof-read the galleys and pinpoint all mistakes in the editing process.

Statement of integrity. The submission of a manuscript conveys the authors' assurance that (1) it has not been published elsewhere before, (2) it has not been submitted to other journals or editorials, (3) its publication has been approved by all co-authors, and (4) permission has been granted to publish all copyrighted material contained within.

All correspondence should be addressed to

PRO MATHEMATICA
Pontificia Universidad Católica del Perú
Departamento de Ciencias
Apartado 1761
Lima 100, Perú

e.mail: promathematica@pucp.edu.pe