Departamento de Cs. e Ingeniería de la Computación

Universidad Nacional del Sur

Bases de Datos

Segundo Cuatrimestre de 2016

Proyección de dependencias funcionales y cálculo de formas normales: ejemplos

1. Proyección de dependencias funcionales

Durante el cálculo de las formas normales una operación muy frecuente es la de calcular la proyección de un conjunto de d.f. F sobre un conjunto de atributos (o subesquema) S, esto es $\Pi_S(F)$. Formalmente $\Pi_S(F)$ se define como el conunto de d.f. de la forma $X \to Y$ tal que $XY \subseteq S$ y $X \to Y \in F^+$ (*i.e.*, $F \models X \to Y$)

A continuación, se mostrará a través de un ejemplo como calcular $\Pi_S(F)$ sin necesidad de calcular F^+.

Sea $F = \{KMS \to TN, LP \to MTS, LT \to K, LN \to S, MT \to L, S \to L, KT \to L\}$ un *conjunto mínimo reducido* y sea $KMST$ un conjunto de atributos.

Pasos para calcular $\Pi_{KMST}(F)$

- Abrir a derecha las d.f. de F: reemplazar cada d.f. en F de la forma $X \to A_1A_2 \dots A_n$ por $X \to A_1$, $X \to A_2$, ..., $X \to A_n$,

  $F = \{KMS \to T, KMS \to N, LP \to M, LP \to T, LP \to S,$
  $LT \to K, LN \to S, MT \to L, S \to L, KT \to L\}$

- Proyectamos las d.f. que aparecen explícitamente en F formadas por atributos que pertenecen a $KMST$ en un conjunto C al cual le iremos incorporando d.f. hasta llegar a $C = \Pi_{KMST}(F)$. En este caso $C = \{KMS \to T\}$,

- Solo restan identificar aquellas d.f. formadas por atributos en $KMST$ que pertenecen a F^+ y no aparecen explícitamente en F.

  Para esto identificamos un conjunto de atributos $Izq_{(KMST)}$ que contiene aquellos atributos de $KMST$ que aparecen en el lado izquierdo de alguna d.f. en F. En este caso $Izq_{(KMST)} = KMST$.

  El objetivo del conjunto $Izq_{(KMST)}$ es identificar aquellos atributos que pueden determinar a otros (*i.e.*, aparecen del lado izquierdo de alguna d.f.) y descartar aquellos atributos que no pueden determinar a otros (*i.e.*, aparecen solo del lado derecho), para facilitar la búsqueda de las d.f. que aparecen en F^+. En este ejemplo no se descarta ningún atributo, pero en general podrían descartarse.

  Luego calcularemos las clausuras de distintas combinaciones de atributos presentes en $Izq_{(KMST)} = KMST$. Podemos descartar las combinaciones de 4 (todos los atributos) dado que estamos buscando d.f. que no sean triviales o redundantes.

  combinaciones de 1 atributo de KMST:

  - $K_F^+ = K \Rightarrow$ no se deduce ninguna d.f. no trivial
  - $M_F^+ = M \Rightarrow$ no se deduce ninguna d.f. no trivial
  - $S_F^+ = SL \Rightarrow$ se deduce $S \to L$, pero no se proyecta porque $L \notin KMST$
  - $T_F^+ = T \Rightarrow$ no se deduce ninguna d.f. no trivial

  combinaciones de 2 atributos de KMST:

  - $KM_F^+ = KM \Rightarrow$ no se deduce ninguna d.f. no trivial
  - $KS_F^+ = KSL \Rightarrow$ se deduce $KS \to L$, pero no se proyecta porque $L \notin KMST$
  - $KT_F^+ = KTL \Rightarrow$ se deduce $KT \to L$, pero no se proyecta porque $L \notin KMST$

- $MS_F^+ = MSL \Rightarrow$ se deduce $MS \to L$, pero no se proyecta porque $L \notin KMST$
- $MT_F^+ = MTLK \Rightarrow$ se deduce que $MT \to K$ **se proyecta sobre** $KMST$, por lo tanto lo incorporamos a C: $C = \{KMS \to T, MT \to K\}$
- $ST_F^+ = STLK \Rightarrow$ se deduce que $ST \to K$ **se proyecta sobre** $KMST$, por lo tanto la incorporamos a C: $C = \{KMS \to T, MT \to K, ST \to K\}$

combinaciones de 3 atributos de KMST:

- $KMS_F^+ = KMSTNL \Rightarrow$ se deduce $KMS \to T$, no es necesario agregarla porque ya pertenece a $C = \{KMS \to T, MT \to K, ST \to K\}$
- $KMT_F^+ = KMTL \Rightarrow$ se deduce $KMT \to L$, pero no se proyecta porque $L \notin KMST$
- $KST_F^+ = KSTL \Rightarrow$ se deduce $KST \to L$, pero no se proyecta porque $L \notin KMST$
- $MST_F^+ = MSTLKN \Rightarrow$ se deduce $MST \to K$, pero no se incorpora a C por ser redundante en C ya que $K \in MST_C^+ = MSTK$ (note que $ST \to K \in C$)

Luego las dependencias de F que se proyectan en $KMST$ son:

$$\Pi_{KMST}(F) = \{KMS \to T, MT \to K, ST \to K\}$$

2. Formas normales

Sea $F = \{KMS \to TN, LP \to MTS, LT \to K, LN \to S, MT \to L, S \to L, KT \to L\}$ un *conjunto mínimo reducido* definido sobre el esquema $R(KLMNPSTU)$ donde las laves candidatas son: $PUL, PUS, PUKT, PUMT$.

2.1. Encontrar una descomposición en 3FN, join sin perdida (j.s.p.), que preserve dependencias (p.d.) y optimizada.

Pasos para obtener una **descomposición en 3FN, j.d.p., p.d., optimizada** de un esquema R con un conjunto d.f. F *mínimo reducido*:

1. Abrir a derecha las dependencias funcionales: reemplazar cada d.f. en F de la forma $X \to A_1A_2 \dots A_n$ por $X \to A_1$, $X \to A_2$, $\dots$ $X \to A_n$,

2. verificar si el esquema R sin descomponer ya se encuentra en 3FN. Para que esto de cumpla cada d.f. $X \to A \in F$ debe satisfacer al menos una de las siguientes condiciones:

   - X es superllave, o
   - A es primo (es parte de alguna llave)

   Nota: *para verificar esto es necesario conocer (o calcular) todas las llaves de candidatas.*

   Si R está en 3FN no es necesario hacer la descomposición (paso **3**).

3. Descomposición: Para cada d.f. $X \to A \in F$ formar un esquema XA y calcular $\Pi_{(XA)}F$ (las dependencias de F que se proyectan sobre el esquema XA).
   Atención! En un esquema se pueden proyectar más de una d.f.. Por lo tanto se formará un esquema para $X \to A \in F$ **siempre y cuando no exista** otra d.f. $Y \to B \in F$ talque $XA \subset YB$.

4. Join Sin Perdida: Si ningún esquema contiene una llave, se agrega a la descomposición un esquema formado por los atributos de alguna llave candidata.

5. Optimización: Se unen los esquemas que comparten al menos una llave.

6. Unir esquemas que estén contenidos uno dentro del otro.

Ejemplo 2.1.1. Consideremos el esquema $R(KLMNPSTU)$ con llaves: $PUL, PUS, PUKT, PUMT$ y el *conj. mínimo reducido* $F = \{KMS \to TN, LP \to MTS, LT \to K, LN \to S, MT \to L, S \to L, KT \to L\}$.

Cálculo de una descomposición 3FN, j..s.p, p.d. y optimizada para $R(KLMNPSTU)$ (ver figura 1):

1. Abrir a derecha las dependencias funcionales de F:

$F = \{KMS \to T, KMS \to N, LP \to M, LP \to T, LP \to S,$
$LT \to K, LN \to S, MT \to L, S \to L, KT \to L\}$

2. verificar si el esquema R sin descomponer ya se encuentra en 3FN.

La d.f. $KMS \to N \in F$ no verifica la 3FN ya que KMS no es superllave y N no es primo. Por lo tanto R no esta en 3FN y es necesario hacer la descomposición (paso **3**)

3. Descomposición:

- Subesq. 1: $(KMST)$ - Dependencias: $\Pi_{KMST}(F) = \{KMS \to T, MT \to K, ST \to K\}$ - llaves: KMS, MST. (*Note que* $MT \to K$ *y* $ST \to K$ *no aparecen explicitamente en* F *pero pertenecen a* F^+ *y por lo tanto se proyectan sobre* $KMST$ *- ver cáculo de* $\Pi_{KMST}(F)$ *en la sección 1*).
- Subesq. 2: $(KMNS)$ - Dependencias: $\Pi^F_{KMSN} = \{KMS \to N\}$ - Llaves: KMS
- Subesq. 3: (LMP) - Dependencias: $\Pi^F_{LMP} = \{LP \to M\}$ - Llaves: LP
- Subesq. 4: (LPT) - Dependencias: $\Pi^F_{LPT} = \{LP \to T\}$ - Llaves: LP
- Subesq. 5: (LPS) - Dependencias: $\Pi^F_{LPS} = \{LP \to S, S \to L\}$ - Llaves: LP, PS
- Subesq. 6: (KLT) - Dependencias: $\Pi^F_{KLT} = \{KT \to L, LT \to K\}$ - Llaves: KT, LT
- Subesq. 7: (LMT) - Dependencias: $\Pi^F_{LMT} = \{MT \to L\}$ - Llaves: MT
- Subesq. 8: (LNS) - Dependencias: $\Pi^F_{LNS} = \{S \to L, LN \to S\}$ - Llaves:LN, NS

Nota: *Si bien* $S \to L \in F$, **no se crea un esquema** SL *dado que* $S \to L$ *se proyecta sobre el esquema* LPS *(creado por la d.f.* $LP \to S$*) y también sobre el esquema* LNS *(creado por* $LN \to S$*).*

4. Join sin Perdida: dado que ningún subesquema contiene una llave de R $(PUL, PUS, PUKT, PUMT)$ se agrega un subesquema formado por los atributos de alguna llave a elección, por ejemplo PUS:

- Subesq. 9: (PUS) - Dependencias: $\Pi^F_{PUS} = \{\}$ - Llaves: PUS

Nota: *sobre el esquema* PUS *no se proyectará ninguna dependencia funcional, dato que* PUS *es una llave y por lo tanto los atributos que la forman no se pueden determinar entre sí. Esto ocurre siempre independientemente de la llave que se elija.*

5. Optimización: Se unen los esquemas que comparten al menos una llave

- Subesq. 1 $\cup$ Subesq. 2: $(KMNST)$ -
  Dependencias: $\Pi^F_{KMST} \cup \Pi^F_{KMNS} = \{KMS \to T, MT \to K, KMS \to N\}$ - llaves: KMS, MST
- Subesq. 3 $\cup$ Subesq. 4 $\cup$ Subesq. 5: $(LMPST)$ -
  Dependencias: $\Pi^F_{LMP} \cup \Pi^F_{LPT} \cup \Pi^F_{LPS} = \{LP \to MST, S \to L\}$ - Llaves: LP, PS
- Subesq. 6: (KLT) - Dependencias: $\Pi^F_{KLT} = \{KT \to L, LT \to K\}$ - Llaves: KT, LT
- Subesq. 7: (LMT) - Dependencias: $\Pi^F_{LMT} = \{MT \to L\}$ - Llaves: MT
- Subesq. 8: (LNS) - Dependencias: $\Pi^F_{LNS} = \{S \to L, LN \to S\}$ - Llaves: LN, NS
- Subesq. 9: (PUS) - Dependencias: $\Pi^F_{PUS} = \{\}$ - Llaves: PUS

6. Unir esquemas que estén contenidos uno dentro del otro: el subesquema (LMT) esta contenido en el suesquema $(LMPST)$ por lo tanto se unen para formar un solo subesquema $(LMPST)$ con dependencias $\{LP \to MST, S \to L, MT \to L\}$ y Llaves: LP, PS, MPT.

Finalmente se obtiene la siguiente **descomposición en 3FN, j.s.p, p.d., optimizada:**
$\rho = (KMNST, KLT, LMPST, LNS, PUS)$

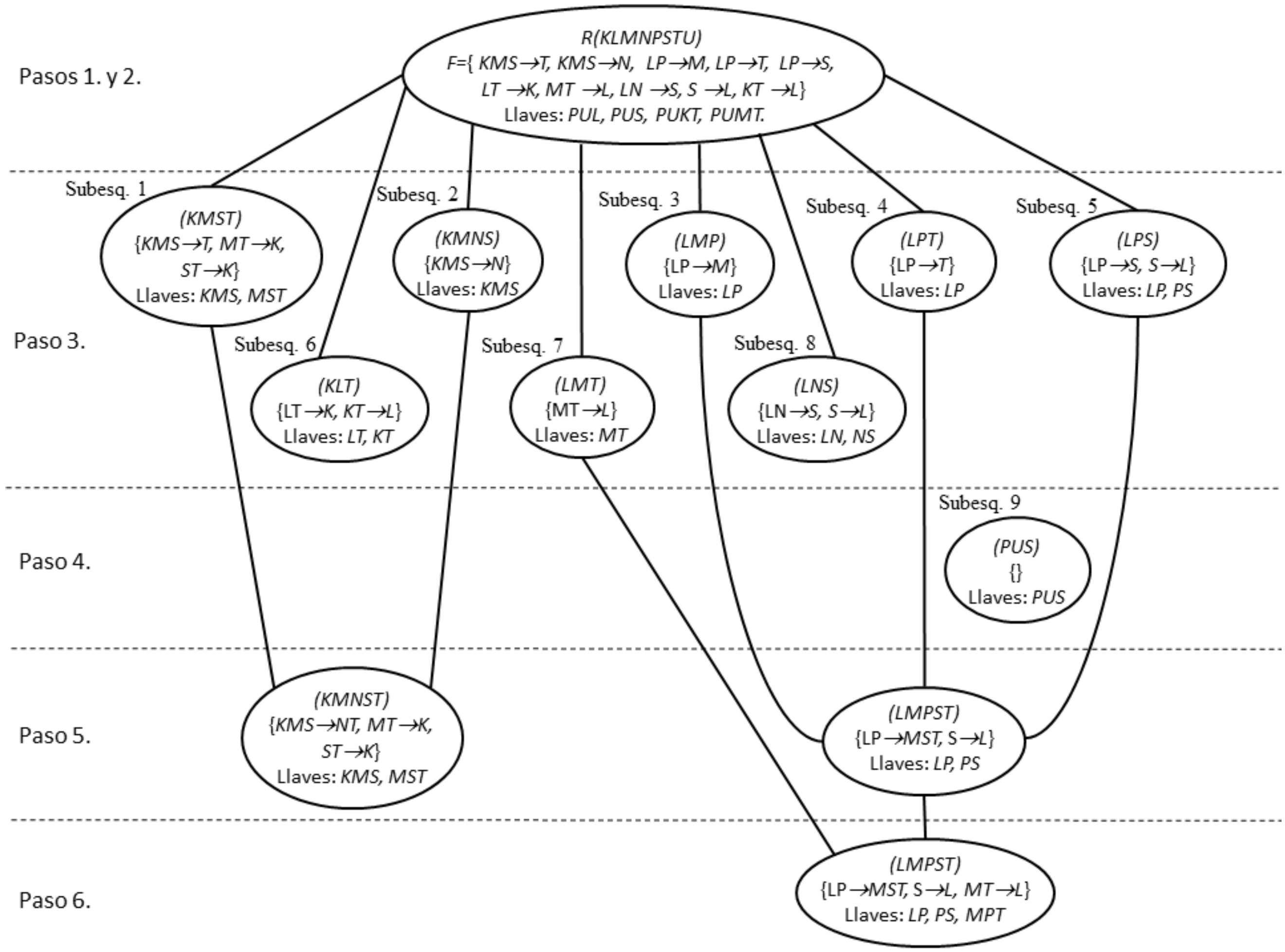

Figura 1: Cálculo de la descomposición $\rho = (KMNST, KLT, LMPST, LNS, PUS)$ en 3FN, p.d., j.s.p., optimizada

2.2. Encontrar una descomposición en FNBC, j.s.p., optimizada.

Pasos a seguir para obtener una **descomposición en FNBC a partir de 3FN p.d.**, j.s.p. de un esquema R con un conjunto de d.f. F:

1. Calcular una descomposición ρ en 3FN, j.s.p., p.d.: Pasos **1**, **2**, **3** y **4** del algoritmo 3FN, j.s.p., p.d.

2. Descomposición de subesquemas que **no** respetan FNBC. Mientras exista un subesquema U en ρ, que viola FNBC:.

a) Sea $X \rightarrow A$ una d.f. de U que viola FNBC, esto es, X no es superllave del subesquema U. Note que A es un solo atributo, por lo tanto las d.f. de U deben estar abiertas a derecha.

Heurística de selección: Si hay mas de una d.f. $X \rightarrow A$ que viola FNBC, elegir primero aquella d.f. con menor cantidad de atributos del lado izquierdo X, y que su atributo A del lado de derecho aparezca en la menor cantidad de d.f. Esto garantiza que se pierdan la menor cantidad de d.f. posibles durante la descomposición.

b) Partir U en dos sub-esquemas XA y U/A, siempre y cuando XA respete F.N.B.C. Para esto es necesario calcular las d.f. que se proyectan en cada sub-esquema.

3. Optimización: Se unen los esquemas que comparten al menos una llave y/o estén contenidos uno dentro del otro **siempre y cuando el esquema resultante respete F.N.B.C.**

Ejemplo 2.2.1. Consideremos el esquema $R(KLMNPSTU)$ con llaves: $PUL, PUS, PUKT, PUMT$ y el *conj. mínimo reducido* $F = \{KMS \to TN, LP \to MTS, LT \to K, LN \to S, MT \to L, S \to L, KT \to L\}$.

Cálculo de una descomposición FNBC, j..s.p y optimizada para $R(KLMNPSTU)$ (ver figura 2):

1. Calculamos una descomposición 3FN, p.d., j.s.p. (pasos **1**, **2**, **3** y **4** del ejemplo anterior)

- Subesq. 1: $(KMST)$ - Dependencias: $\Pi_{KMST}(F) = \{KMS \to T, MT \to K, ST \to K\}$ - llaves: KMS, MTS.
- Subesq. 2: $(KMNS)$ - Dependencias: $\Pi^F_{KMSN} = \{KMS \to N\}$ - Llaves: KMS
- Subesq. 3: (LMP) - Dependencias: $\Pi^F_{LMP} = \{LP \to M\}$ - Llaves: LP
- Subesq. 4: (LPT) - Dependencias: $\Pi^F_{LPT} = \{LP \to T\}$ - Llaves: LP
- Subesq. 5: (LPS) - Dependencias: $\Pi^F_{LPS} = \{LP \to S, S \to L\}$ - Llaves: LP, PS
- Subesq. 6: (KLT) - Dependencias: $\Pi^F_{KLT} = \{KT \to L, LT \to K\}$ - Llaves: KT, LT
- Subesq. 7: (LMT) - Dependencias: $\Pi^F_{LMT} = \{MT \to L\}$ - Llaves: MT
- Subesq. 8: (LNS) - Dependencias: $\Pi^F_{LNS} = \{S \to L, LN \to S\}$ - Llaves:LN, NS
- Subesq. 9: (PUS) - Dependencias: $\Pi^F_{PUS} = \{\}$ - Llaves: PUS

2. Descomposición de subesquemas que no respetan FNBC (sombreados con gris en la figura 2):

- el Subesq. 1 $(KMST)$ con llave: KMS no respeta F.N.B.C. por que $\Pi_{KMST}(F)$ contiene la dependencia $ST \to K$ y ST no es superllave. Luego $(KMST)$ se descompone en dos subesquemas:
  - Subesquema 1.1: (KST) - Dependencias: $\Pi_{KST}(\Pi_{KMST}(F)) = \{ST \to K\}$ llaves: ST
  - Subesquema 1.2: (MST) - Dependencias: $\Pi_{MST}(\Pi_{KMST}(F)) = \{\}$ llaves: MST
    Note que las d.f. que se proyectan en cada subesquema provienen de las dependencias proyectadas en $KMST$ ($\Pi_{KMST}(F)$) y no de F.
- el Subesq. 5 (LPS) con llaves: LP, PS, no respeta F.N.B.C. por que $\Pi_{LPS}(F)$ contiene la dependencia $S \to L$ y S no es superllave. Luego (LPS) se descompone en dos subesquemas:
  - Subesquema 5.1: (SL) - Dependencias: $\Pi_{SL}(\Pi_{LPS}) = \{S \to L\}$ llaves: S
  - Subesquema 5.2: (PS) - Dependencias: $\Pi_{PS}(\Pi_{LPS}) = \{\}$ llaves: PS
- el Subesq. 8 (LNS) con llaves: LN, NS, no respeta F.N.B.C. por que $\Pi_{LNS}(F)$ contiene la dependencia $S \to L$ y S no es superllave. Luego (LNS) se descompone en dos subesquemas:
  - Subesquema 8.1: (SL) - Dependencias: $\Pi_{SL}(\Pi_{LNS}) = \{S \to L\}$ llaves: S
  - Subesquema 8.2: (NS) - Dependencias: $\Pi_{NS}(\Pi_{LNS}) = \{\}$ llaves: NS

Los subesquemas 2,3,4,6,7 y 9 respetan FNBC y no hay que descomponerlos.

3. Optimización:

- Los subesquemas 3 (LMP) y 4(LPT) comparten la llave LP. Al unirlos, se genera el subesquema (LMPT) que contiene al subesquema 7(LMT), quedando:
  $(LMPT)$ con dependencias: $\Pi^F_{LMPT} = \{LP \to MT, MT \to L\}$ y llaves:LP, MPT
  La dependencia $MT \to L$ viola FNBC porque MT no es superllave, por lo tanto no se puede hacer esta optimización.
- Los subesquemas 8.1(LS) y 5.1 (LS) comparten la llave y están contenidos uno dentro del otro (son iguales) por lo tanto se funden en uno solo.
- El subesquema 8.2 (NS) esta contenido dentro del subesquema 2 $(KMNS)$, por lo tanto se funden en un solo esquema $(KMNS)$ con las dependencias: $\{KMS \to N\}$ y llave KMS.
- El subesquema 5.2 (PS) esta contenido dentro del subesquema 9 (PUS), por lo tanto se funden en un solo esquema (PUS) sin dependencias y llave PUS.

Finalmente se obtiene la siguiente **descomposición en FNBC, j.s.p, optimizada:**
$\rho = (KLT, KST, MST, KMSN, LS, LMT, LMP, LPT, PUS)$

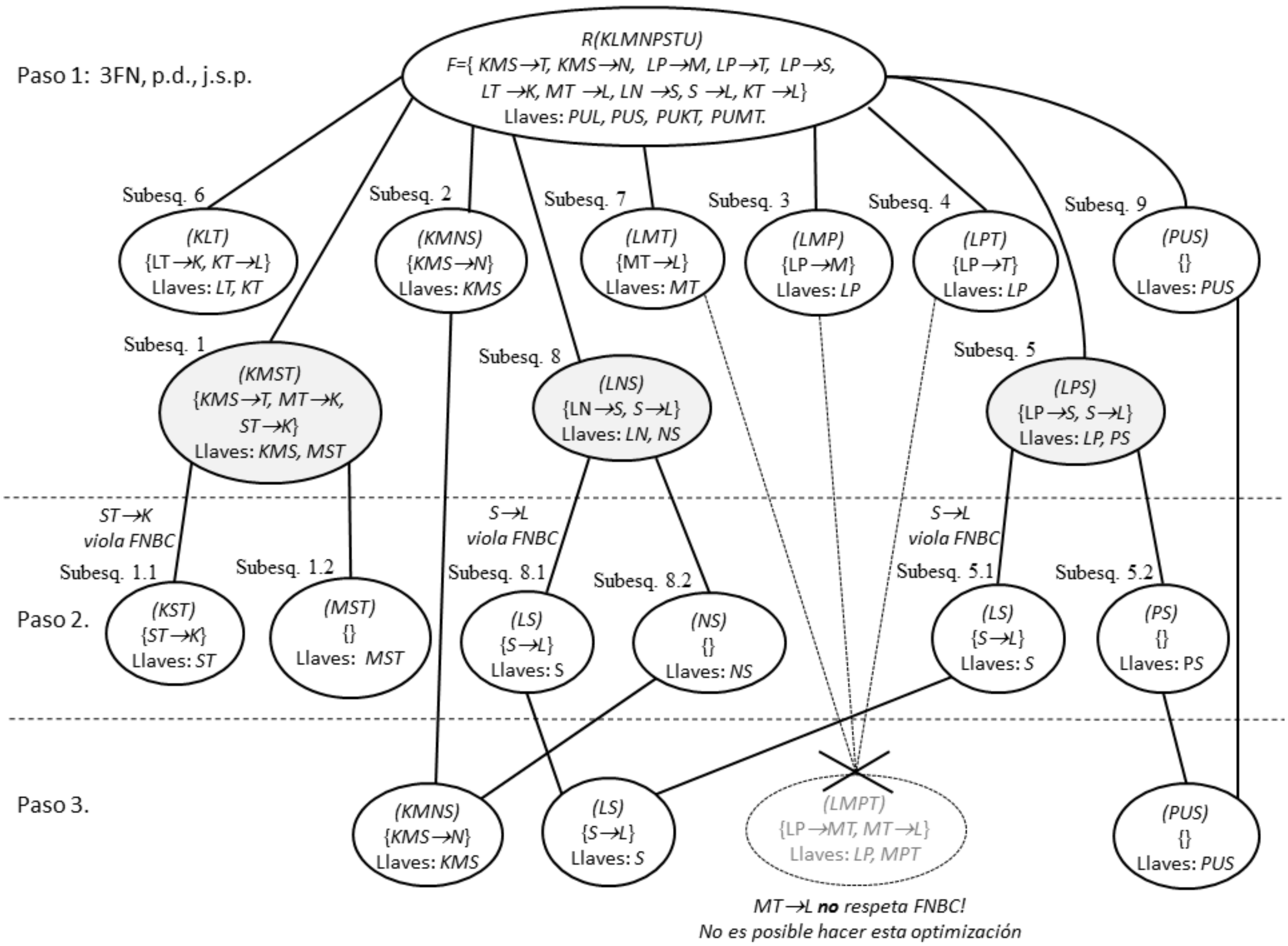

Figura 2: Cálculo de la descomposición $\rho = (KLT, KST, MST, KMSN, LS, LMT, LMP, LPT, PUS)$ en FNBC, j.s.p., optimizada.