Análisis y diseño de algoritmos

7. Ramificación y Poda

José Luis Verdú Mas, Jose Oncina, Mikel L. Forcada

Dep. Lenguajes y Sistemas Informáticos
Universidad de Alicante

30-01-2018 (389)

Índice

Ejemplo Introductorio

El problema de la mochila (general)

Dados:

- n objetos con valores v_i y pesos w_i
- una mochila que solo aguanta un peso máximo W

Seleccionar un conjunto de objetos de forma que:

- no se sobrepase el peso límite W (restricción)
- el valor transportado sea máximo (función objetivo)

Formalización

- Solución: $X = (x_1, x_2, \ldots, x_n) \quad x_i \in \{0, 1\}$
- Restricciones:
  - Implícitas:

$$x_i \in \begin{cases} 0 & \text{no se selecciona el objeto } i \\ 1 & \text{se selecciona el objeto } i \end{cases}$$

  - Explícitas:

$$\sum_{i=1}^{n} x_i w_i \leq W$$

- Función objetivo:

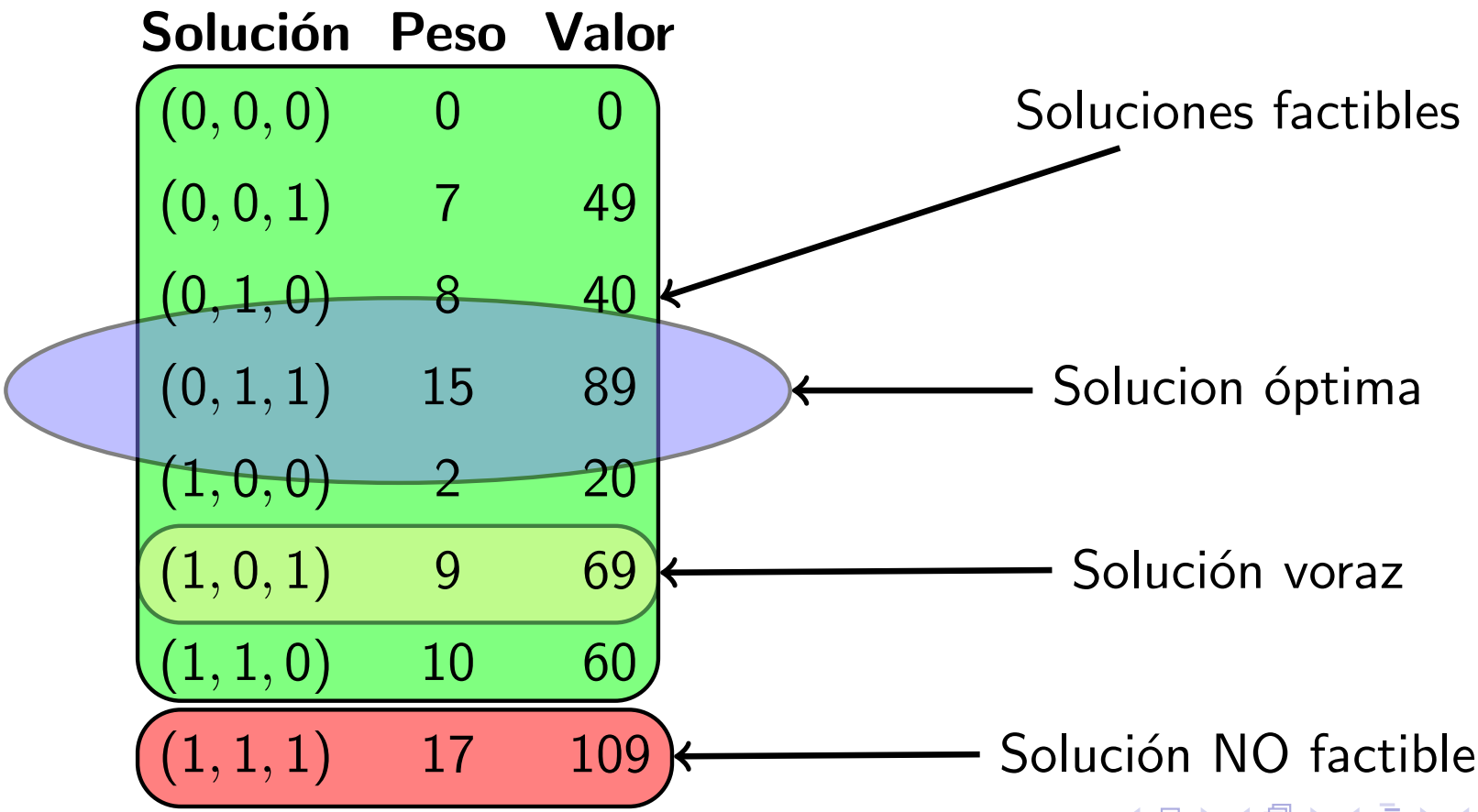
$$\text{máx} \sum_{i=1}^{n} x_i v_i$$

Tipos de soluciones

- Supongamos el siguiente ejemplo:

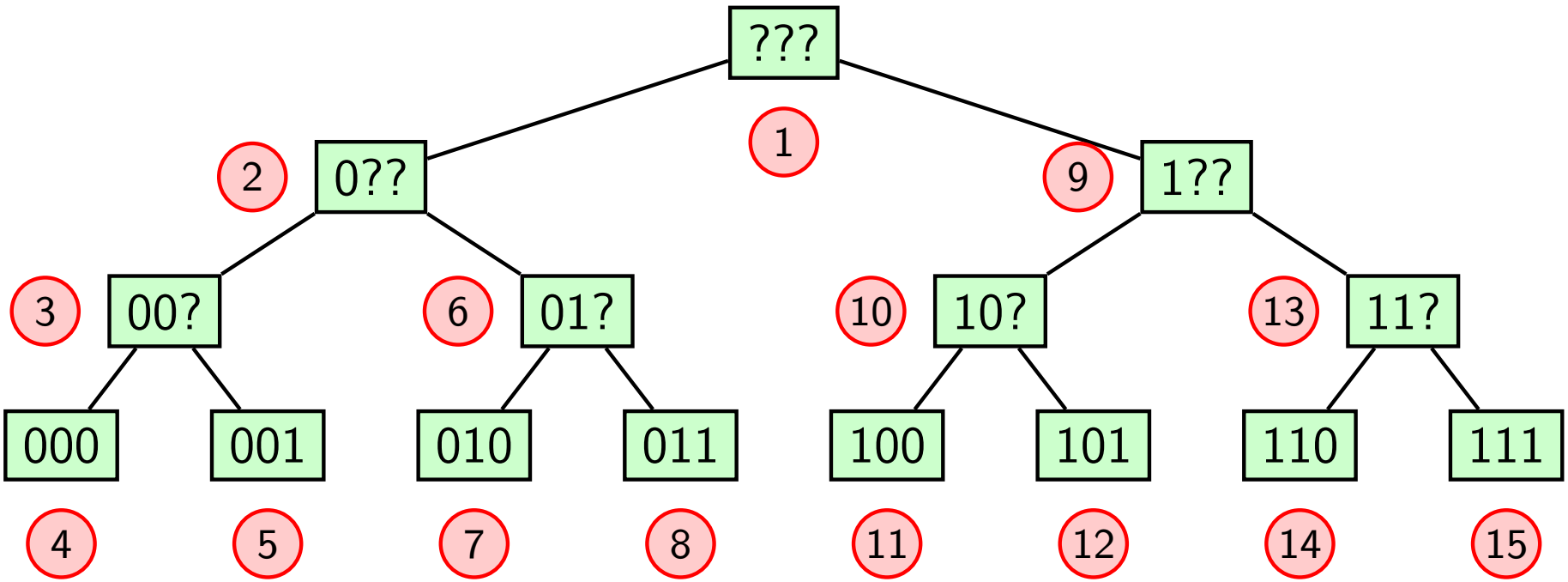
$$W = 16 \qquad w = (2, 8, 7) \qquad v = (20, 40, 49)$$

- Combinaciones posibles (espacio de soluciones):



Solución usando vuelta atrás

- Combinaciones posibles:
  - Supongamos el ejemplo: $W = 16 \quad w = (7, 8, 2) \quad v = (49, 40, 20)$
  - Espacio de soluciones
    - Generación ordenada mediante vuelta atrás
    - Nodos generados: 15
    - Nodos expandidos: 7



¿Es el recorrido importante?

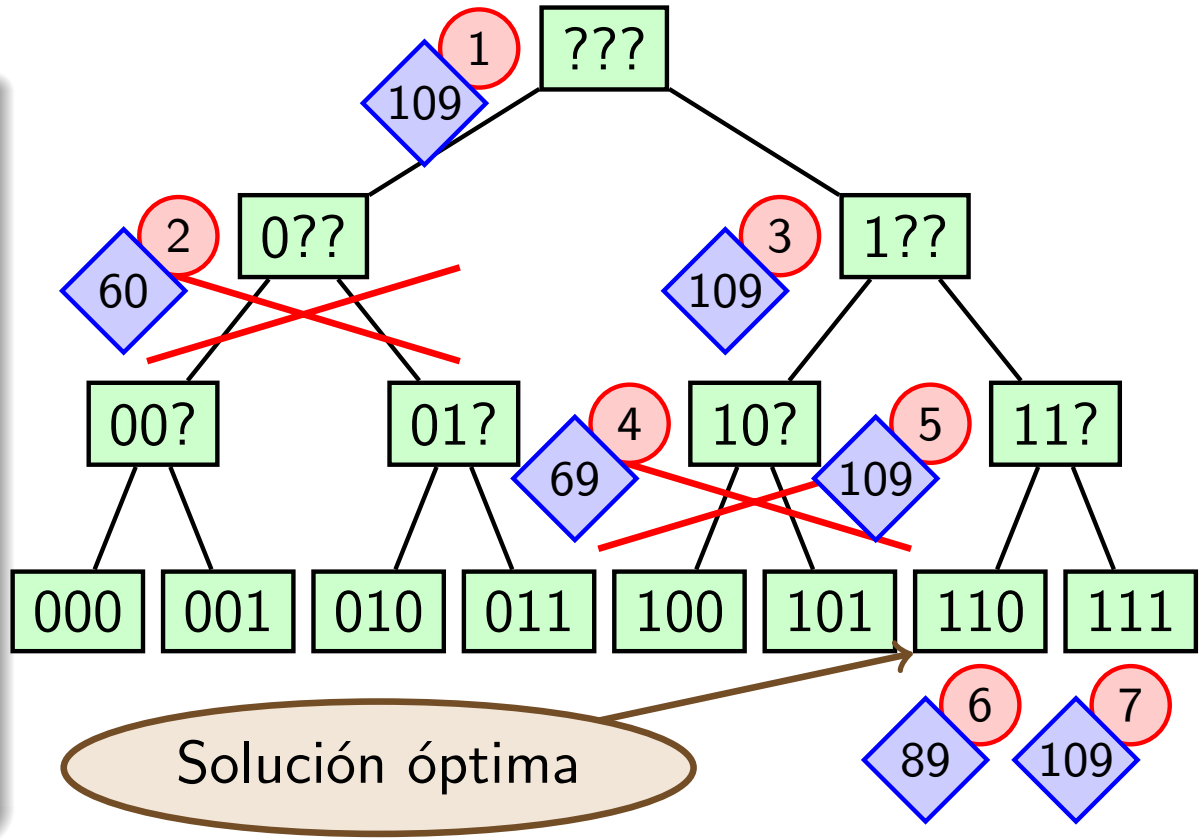
- ¿Podríamos llegar antes a la solución óptima con otro recorrido?
- ¿Cómo?
  - Adecuando el orden de exploración del árbol de soluciones según nuestros intereses. Se priorizará para su exploración aquellos nodo mas prometedores
- ... sin olvidar las podas

Ejemplo Introductorio

- Combinaciones posibles:
  - Supongamos el ejemplo: $W = 16 \quad w = (7, 8, 2) \quad v = (49, 40, 20)$

Función de cota

Cada nodo toma cota el valor que resultaría de incluir en la solución aquellos objetos pendientes de tratar (sustituir en cada nodo los '?' por '1') independientemente de que quepan o no.

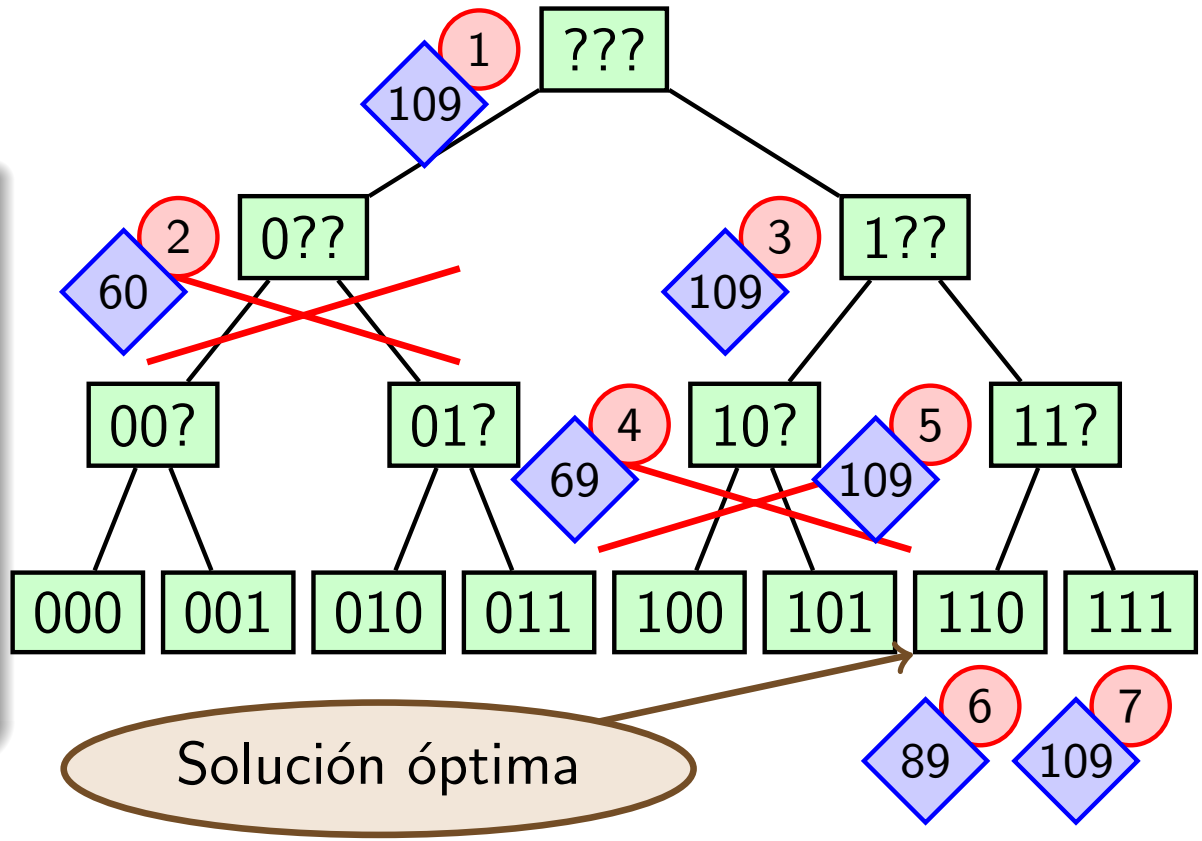


Ejemplo Introductorio

- Combinaciones posibles:
  - Supongamos el ejemplo: $W = 16 \quad w = (7, 8, 2) \quad v = (49, 40, 20)$

Espacio de soluciones

- Orden de expansión priorizando los nodos con mayor cota
- Generados: 7
- Expandidos: 3
- Reducción $\geqslant 50\,\%$



Implementación

```
1  float knapsack( const vector<double> &v, const vector<double> &w, double W ) {
2
3      typedef vector<short> sol;
4      typedef tuple<double, double, sol, int > node; // value, weight, vector, k
5      priority_queue< node > pq;
6
7      double best_val = knapsack_d( v, w, 0, W);
8      pq.push( node(0.0, 0.0, sol(v.size()), 0 ) );
9
10     while( !pq.empty() ) {
11         double value, weight;
12         sol x;
13         unsigned k;
14
15         tie(value, weight, x, k) = pq.top();
16         pq.pop();
17         /* ... next slide ...*/
18     }
19     return best_val;
20 }
```

Dentro del bucle ...

```
1  /* ... */
2      if( k == v.size() ) {                              // base case
3          best_val = max( value, best_val ) ;
4          continue;
5      }
6
7      for (unsigned j = 0; j < 2; j++ ) {              // no base case
8          x[k] = j;
9
10         double new_weight = weight + x[k] * w[k]; // updating weight
11         double new_value = value + x[k] * v[k];     // updating value
12
13         if( new_weight <= W &&                        // is feasible & is promising
14             new_value + knapsack_c( v, w, k + 1, W - new_weight ) > best_val
15         )
16             pq.push( node( new_value, new_weight, x, k + 1 ) );
17     }
18 /* ... */
```

Usando podas pesimistas

```
1  /* ... */
2      if( k == v.size() ) {                                    // base case
3          best_val = max( value, best_val ) ;
4          continue;
5      }
6
7      for (unsigned j = 0; j < 2; j++ ) {
8          x[k] = j;
9
10         double new_weight = weight + x[k] * w[k];          // updating weight
11         double new_value = value + x[k] * v[k];             // updating value
12
13         if( new_weight <= W ) {                                  // is feasible
14                                                  // updating pessimistic bound
15             best_val = max( best_val, new_value
16                 + knapsack_d( v, w, k+1, W - new_weight )
17             );
18
19             double opt_bound = new_value + knapsack_c(v, w, k+1, W-new_weight);
20             if( opt_bound > best_val )                          // is promising
21                 pq.push( node( new_value, new_weight, x, k+1 ) );
22         }
23     }
24 /* ... */
```

Ordenando por cota optimista

```
1  double knapsack( const vector<double> &v, const vector<double> &w, double W ) {
2    typedef vector<short> sol;
3    //              opt_bound, value,  weight,   x,   k
4    typedef tuple< double, double, double, sol, unsigned > node;
5    priority_queue< node > pq;
6
7    double best_val = knapsack_d( v, w, 0, W);
8    double opt_bound = knapsack_c(v, w, 0, W);
9
10   pq.push( node( opt_bound, 0.0, 0.0, sol(v.size()), 0 ) );
11
12   while( !pq.empty() ) {
13     double value, weight;
14     sol x;
15     unsigned k;
16
17     tie(ignore, value, weight, x, k) = pq.top();
18     pq.pop();
19     /* ... Next slide ... */
20   }
21   return best_val;
22 }
```

Dentro del bucle

```
1  /* ... */
2      if( k == v.size() ) {  //  base case
3          best_val = max( best_val, value ) ;
4          continue;
5      }
6
7      for (unsigned j = 0; j < 2; j++ ) {  // non base
8          x[k] = j;
9
10         float new_weight = weight + x[k] * w[k]; // updating weight
11         float new_value = value + x[k] * v[k];   // updating value
12
13         if( new_weight <= W ) {
14             best_val = max( best_val, new_value
15                 + knapsack_d( v, w, k+1, W - new_weight));
16             float opt_bound =  new_value
17                 + knapsack_c( v, w, k+1, W - new_weight);
18             if( opt_bound > best_val)  // is promising
19                 pq.push( node( opt_bound, new_value, new_weight, x, k+1 ) );
20         }
21     }
22 /* ... */
```

Comparación

Promedio del número de iteraciones para 100 instancias aleatorias del problema de la mochila con 100 objetos.

	optimista	inicializando	pesimista
Vuelta Atrás	4 491	277	253
RyP (por valor)	2 406	229	197
RyP (por cota optimista)	206	122	112

Índice

Definición y ámbito de aplicación

- Variante del diseño de Vuelta Atrás
  - Realiza una enumeración parcial del espacio de soluciones mediante la generación de un árbol de expansión
  - Uso de cotas para podar ramas que no son de interés
- Nodo vivo: aquel con posibilidades de ser ramificado (visitado pero no completamente expandido)
- Los nodos vivos se almacenan en estructuras que faciliten su recorrido y eficiencia de la búsqueda:

  - En profundidad (estrategia LIFO) $\Rightarrow$ pila
  - En anchura (estrategia FIFO) $\Rightarrow$ cola
  - Dirigida (primero el mas prometedor) $\Rightarrow$ cola de prioridad

Definición y ámbito de aplicación

- Funcionamiento de un algoritmo de ramificación y poda
- Etapas
  - Partimos del nodo inicial del árbol
  - Se asigna una solución pesimista (subóptima, soluciones voraces)
  - Selección
    - Extracción del nodo a expandir del conjunto de nodos vivos
    - La elección depende de la estrategia empleada
    - Se actualiza la mejor solución con las nuevas soluciones encontradas
  - Ramificación
    - Se expande el nodo seleccionado en la etapa anterior dando lugar al conjunto de sus nodos hijos
  - Poda
    - Se eliminan (podan) nodos que no contribuyen a la solución
    - El resto de nodos se añaden al conjunto de nodos vivos
  - El algoritmo finaliza cuando se agota el conjuntos de nodos vivos

Proceso de poda

- Cota optimista:
  - estima, a mejor, el mejor valor que podría alcanzarse al expandir el nodo
  - puede que no haya ninguna solución factible que alcance ese valor
  - normalmente se obtienen relajando las restricciones del problema
  - si la cota optimista de un nodo es peor que la solución en curso, se puede podar el nodo
- Cota pesimista:
  - estima, a peor, el mejor valor que podría alcanzarse al expandir el nodo
  - ha de asegurar que existe una solución factible con un valor mejor que la cota
  - normalmente se obtienen mediante soluciones voraces del problema
  - se puede eliminar un nodo si su cota optimista es peor que la mejor cota pesimista
  - permite la poda aún antes de haber encontrado una solución factible
- Cuanto mas ajustadas sean las cotas, mas podas se producirán

Esquema de Ramificación y Poda

```
1  solution branch_and_bound( problem p ) {
2
3      node initial = initial_node(p);                    // supossed feasible
4      solution current_best = pessimistic_solution(initial);   // pessimistic
5      priority_queue<Node> q.push(initial);
6
7      while( ! q.empty() ) {
8          node n = q.top();
9          q.pop();
10
11         if( is_leaf(n) ) {
12             if( is_better( solution(n), current_best) )
13                 current_best = solution(n);
14             continue;
15         }
16
17         for( node a : expand(n) )
18             if( is_feasible(a) && is_promising( a, current_best))
19                 q.push(a);
20     }
21
22     return current_best;
23 }
```

Esquema de Ramificación y Poda

- Funciones:
  - `initial_node(p)`: obtiene el nodo inicial para la expansión
  - `pesimistic_solution(n)`: devuelve una solución aproximada (factible pero no la óptima)
  - `is_leaf(n)`: mira si `n` es una posible solución
  - `solution(n)` devuelve el valor del nodo `n`
  - `expand(n)`: devuelve la expansión de `n`
  - `is_feasible(n)`: comprueba si `n` cumple las restricciones
  - `is_promising( n, current_best )`: mira si a partir del nodo `n` se pueden obtener soluciones mejores que `current_best` (normalmente se obtiene mediante una solución optimista)

Podando con cotas pesimistas

```
1  solution branch_and_bound( problem p ) {
2      node initial = initial_node(p);                    // supossed feasible
3      solution current_best = pessimistic_solution(initial);  // pessimistic
4      priority_queue<Node> q.push(initial);
5
6      while( ! q.empty() ) {
7          node n = q.top();
8          q.pop();
9
10         if( is_leaf(n) ) {
11             if( is_better( solution(n), current_best) )
12                 current_best = solution(n);
13             continue;
14         }
15         for( node a : expand(n) )
16             if( is_feasible(a) ) {
17                 if( is_better( pessimistic_solution(n), current_best ) )
18                     current_best = pessimistic_solution(n);
19                 if( is_promising (a, current_best ))
20                     q.push(a);
21             }
22     }
23     return current_best;
24   }
```

Características

- La estrategia puede proporcionar:
  - Todas las soluciones factibles
  - Una solución al problema
  - La solución óptima al problema
  - Las n mejores soluciones
- Objetivo de esta técnica
  - Mejorar la eficiencia en la exploración del espacio de soluciones
- Desventajas/Necesidades
  - Encontrar una buena cota optimista (problema relajado)
  - Encontrar una buena solución pesimista (estrategias voraces)
  - Encontrar una buena estrategia de exploración (cómo ordenar)
  - Mayor requerimiento de memoria que los algoritmos de Vuelta Atrás
  - las complejidades en el peor caso suelen ser muy altas
- Ventajas
  - Suelen ser más rápidos que Vuelta Atrás

Esquema de Ramificación y Poda

- Muchas veces, para ver si un nodo es prometedor, se hace comparando la mejor solución obtenida con una solución optimista de nodo.

```
1 bool is_promising( const node &n, const solution &current_best) {
2   return is_better( optimistic_solution(n), current_best );
3 }
```

se pueden hacer cotas agresivas cambiándola por:

```
1 bool is_promising( const node &n, const solution &current_best) {
2   return is_significantly_better(optimistic_solution(n), current_best);
3 }
```

¡Cuidado! puede que se pierda la solución óptima

Índice

El viajante de comercio

Dado un grafo ponderado $g = (V, A)$ con pesos no negativos, el problema consiste en encontrar un *ciclo hamiltoniano* de mínimo coste.

- Un *ciclo hamiltoniano* es un recorrido en el grafo que recorre todos los vértices sólo una vez y regresa al de partida.
- El coste de un ciclo viene dado por la suma de los pesos de las aristas que lo componen.

El viajante de comercio

- Expresamos la solución mediante una tupla $X = (x_1, x_2, \ldots, x_n)$ donde $x_i \in \{1, 2, \ldots, n\}$ es el vértice visitado en i-ésimo lugar.
  - Asumimos que los vértices están numerados, $V = \{1, 2, \ldots, n\}, \quad n = |V|$
  - Fijamos el vértice de partida (para evitar rotaciones):
    - $x_1 = 1;\ x_i \in \{2, 3, \ldots, n\} \quad \forall i : 2 \leq i \leq n$
- Restricciones
  - No se puede visitar dos veces el mismo vértice: $i \neq j \rightarrow x_i \neq x_j \quad \forall i, j : 1 \leq i \leq n \quad 1 \leq j \leq n$
  - Existencia de arista: $\forall i : 1 \leq i < n,\ peso(g, x_i, x_{i+1}) \neq \infty$
  - Existencia de arista que cierra el camino: $peso(g, x_n, x_1) \neq \infty$
- Función objetivo:

$$\min \sum_{i=1}^{n-1} peso(g, x_i, x_{i+1}) + peso(g, x_1, x_n)$$

El viajante de comercio

```
1  unsigned travelling_salesman( const graph &g) {
2      struct node {
3          unsigned opt_bound, length;
4          vector<short> x;
5          unsigned k;
6      };
7      struct is_worse {
8          bool operator() (const Node& a, const Node& b) {
9              return a.opt_bound > b.opt_bound;
10         }
11     };
12     priority_queue< node, vector<node>, is_worse > pq;
13     vector<short> x(g.num_cities());
14     for( unsigned i = 0; i < g.num_cities(); i++ )  x[i] = i;
15     unsigned shortest = pessimistic_bound( g, x, 1);
16     unsigned opt_bound = optimistic_bound( g, x, 1);
17     pq.push( { opt_bound, 0, x, 1 } );
18     while( !pq.empty() ) {
19       node n = pq.top();
20       pq.pop();
21     /* ... Next slide ... */
22     }
23     return shortest;
24 }
```

Dentro del bucle ...

```
1  /* ... */
2      if( n.k == g.num_cities() ) {
3        shortest = min( shortest, n.length + g.dist(n.x[n.k-1],n.x[0]) );
4        continue;
5      }
6      for( unsigned c = n.k; c < n.x.size(); c++ ) {
7        swap( n.x[n.k], n.x[c] );
8
9        unsigned new_length = n.length + g.dist(n.x[n.k-1],n.x[n.k]);
10       unsigned opt_bound = new_length + optimistic_bound(g,n.x,n.k+1);
11       unsigned pes_bound = new_length + pessimistic_bound(g,n.x,n.k+1);
12
13       shortest = min( shortest, pes_bound);
14
15       if( opt_bound <= shortest )
16         pq.push( { opt_bound, new_length, n.x, n.k+1 } );
17
18       swap( n.x[n.k], n.x[c] );
19     }
20 /* ... */
```

Cambiando la prioridad ...

Por cota optimista

```
1 struct is_worse {
2     bool operator() (const node& a, const node& b) {
3         return a.opt_bound > b.opt_bound;
4     }
5 };
```

Por distancia recorrida

```
1 struct is_worse {
2     bool operator() (const node& a, const node& b) {
3         return a.length > b.length;
4     }
5 };
```

Por distancia media recorrida

```
1 struct is_worse {
2     bool operator() (const node& a, const node& b) {
3         return a.length/a.k > b.length/b.k;
4     }
5 };
```

Algunos ejemplos de ejecución

Promedio del número de iteraciones necesitado para resolver 100 instancias del problema del viajante de comeecio con 15 ciudades

- **Cota optimista**: minumum spanning tree de las ciudades restantes
- **Cota pesimista**: algoritmo voraz basado en la ciudad más cercana

Algoritmo	**cota opt.**	**dist. recorrida**	**dist. media**
Vuelta atrás	23 478	23 478	23 478
Ramificación y poda	10 798	12 285	11 421
factor de aceleración	2.17	1.91	2.05

La función compuesta mínima

Dadas dos funciones $f(x)$ y $g(x)$ y dados dos números cualesquiera x e y, encontrar la función compuesta mínima que obtiene el valor y a partir de x tras aplicaciones sucesivas e indistintas de $f(x)$ y $g(x)$

- Ejemplo: Sean $f(x) = 3x$, $g(x) = \lfloor x/2 \rfloor$, y sean $x = 3$, $y = 6$
  - Una transformación de 3 en 6 con operaciones f y g es:

$$(g \circ f \circ g \circ f \circ f \circ g)(3) = 6 \qquad \text{(5 composiciones)}$$

  - La mínima (aunque no única) es:

$$(f \circ g \circ g \circ f)(3) = 6 \qquad \text{(3 composiciones)}$$

La función compuesta mínima

Solución:

- $X = (x_1, x_2, \ldots, x_?) \quad x_i \in \{0, 1\} \begin{cases} 0 \equiv & \text{se aplica } f(x) \\ 1 \equiv & \text{se aplica } g(x) \end{cases}$
  - $(x_1, x_2, \ldots, x_k) \equiv (x_k \circ \cdots \circ x_2 \circ x_1)$
  - El tamaño de la tupla no se conoce a priori
    - Se pretende minimizar el tamaño de la tupla solución (función objetivo)
    - asumiremos un máximo de M composiciones (evitar ramas infinitas)
- Llamanmos $F(X, k, x)$ al resultado de aplicar al valor x la composición representada en la tupla X hasta su posición k

$$F(X, k, x) = (x_k \circ \cdots \circ x_2 \circ x_1)(x) = x_k(\ldots x_2(x_1(x))\ldots)$$

- Restricciones:
  - $F(X, k, x) \neq F(X, i, x)\ \forall i < k$, para recalculos
  - $k < M$, para evitar búsquedas infinitas
  - siempre se puede calcular $F(X, k, x)$
  - $k < v_b$, tupla "prometedora"

Composición de funciones

```
1  int composition( unsigned M, int first, int last ) {
2
3      typedef tuple< short, int> node; // steps, hit
4
5      struct is_worse {
6          bool operator() (const node& a, const node& b) {
7              return get<0>(a) > get<0>(b);
8          }
9      };
10     priority_queue< node, vector<node>, is_worse > pq;
11
12     unsigned best = numeric_limits<unsigned>::max();
13     pq.push( node( 0, first) );
14
15     while( !pq.empty() ) {
16         node n = pq.top();
17         pq.pop();
18     /* ... next slide ... */
19     }
20     return best;
21 }
```

Dentro del bucle ... (sin podas)

```
1   ...
2     unsigned k = get<0>(n);
3     int hit = get<1>(n);
4
5     if( hit == last && k < best )
6         best = k;
7
8     if( k ==  M )
9         continue;
10
11    for( short i = 0; i < 2; i++ ) {
12        int next = F1(i,hit);
13        pq.push( node( k+1, next ) );
14    }
15  ...
```

Dentro del bucle ... (poda: mejor en curso)

poda: mejor en curso

```
1  ...
2      unsigned k = get<0>(n);
3      int hit = get<1>(n);
4
5      if( k >= best )
6          continue;
7
8      if( hit == last && k < best ) {
9          best = k;
10         continue;
11     }
12
13     if( k ==  M )
14         continue;
15
16     for( short i = 0; i < 2; i++ ) {
17         int next = F1(i,hit);
18         pq.push( node( k+1, next ) );
19     }
20 ...
```

Podando con memoria

```
1  int composition( unsigned M, int first, int last ) {
2
3      typedef tuple< short, int> node; // steps, hit
4
5      struct is_worse {
6          bool operator() (const node& a, const node& b) {
7              return get<0>(a) > get<0>(b);
8          }
9      };
10
11     priority_queue< node, vector<node>, is_worse > pq;
12
13     unordered_map<int, Default<unsigned, u_max>> best;
14
15     pq.push( node( 0, first) );
16     while( !pq.empty() ) {
17         node n = pq.top();
18         pq.pop();
19     /* ... next slide ... */
20     }
21     return best[last];
22 }
```

Dentro del bucle ...

```
1  /* ... */
2      unsigned k = get<0>(n);
3      int hit = get<1>(n);
4
5      if( k >= best[last] )
6          continue;
7
8      if( best[hit] <= k )
9          continue;
10     best[hit] = k;
11
12     if( k == M )
13         continue;
14
15     for( short i = 0; i < 2; i++ ) {
16         int next = F1(i,hit);
17         pq.push( Node( k+1, next ) );
18     }
19 /* ... */
```

Ejemplo de ejecución

Número de iteraciones para alcanzar el valor 11 desde 1 (con $M = 20$)

Algoritmo	**Vuelta Atrás**	**Ramificación y Poda**
básico	65 535	65 535
mejor en curso	9 073	8 311
memorizando	565	329

Índice

El Puzzle

- Disponemos de un tablero con n^2 casillas y de $n^2 - 1$ piezas numeradas del uno al $n^2 - 1$. Dada una ordenación inicial de todas las piezas en el tablero, queda sólo una casilla vacía (con valor 0), a la que denominamos *hueco*.
- El objetivo del juego es transformar dicha disposición inicial de las piezas en una disposición final ordenada, en donde en la casilla (i, j) se encuentra la pieza numerada $n(i-1)+j$ y en la casilla (n, n) se encuentra el hueco
- Los únicos movimientos permitidos son los de las piezas adyacentes al hueco (horizontal y verticalmente), que pueden ocuparlo. Al hacerlo, dejan el hueco en la posición en donde se encontraba la pieza antes del movimiento. Otra forma de abordar el problema es considerar que lo que se mueve es el hueco, pudiendo hacerlo hacia arriba, abajo, izquierda o derecha. Al moverse, su casilla es ocupada por la pieza que ocupaba la casilla a donde se ha *movido* el hueco

El Puzzle

- Por ejemplo, para el caso n = 3 se muestra a continuación una disposición inicial junto con la disposición final:

1	5	2
4	3	
7	8	6

Disposición Inicial

1	2	3
4	5	6
7	8	

Disposición final

El Puzzle

- Regla: una pieza se puede mover de A a B si:
  - A está al lado de B
  - B es el hueco

- Según se relaje el problema tenemos dos funciones de coste diferentes:
  1. Calcular el número de piezas que están en una posición distinta de la que les corresponde en la disposición final
  2. Calcular la suma de las distancias de Manhattan desde la posición de cada pieza a su posición en la disposición final
- La distancia de Manhattan entre dos puntos del plano de coordenadas (x_1, y_1) y (x_2, y_2) viene dada por la expresión:

$$|x_1 - x_2| + |y_1 - y_2|$$