Complete set of recursions for the pseudoknot algorithm

Elena Rivas[1] **and Sean R. Eddy**[2]

Department of Genetics,

Washington University, St. Louis, MO 63130 USA

Supplementary material for

"A dynamic programming algorithm for
RNA structure prediction including pseudoknots"
by Elena Rivas and Sean R. Eddy.

[1] Email: elena@genetics.wustl.edu.
[2] Email: eddy@genetics.wustl.edu.

In this supplementary material we provide the complete recursion relations for all the matrices used in the pseudoknot algorithm presented in "A dynamic programming algorithm for RNA structure prediction including pseudoknots" by E. Rivas and S. Eddy. The algorithm is a dynamic programming algorithm for RNA folding including pseudoknots that runs $\mathcal{O}(N^6)$ in time and $\mathcal{O}(N^4)$ in storage.

The description of the five matrices to be filled in the dynamic programming algorithm is given in table 1.

matrix $(i \leq k \leq l \leq j)$	relationship i, j	relationship k, l
$vx(i,j)$	paired	—
$wx(i,j)$	undetermined	—
$vhx(i,j:k,l)$	paired	paired
$zhx(i,j:k,l)$	paired	undetermined
$yhx(i,j:k,l)$	undetermined	paired
$whx(i,j:k,l)$	undetermined	undetermined

Table 1: Specifications of the matrices used in the pseudoknot algorithm.

Conventions include: the $5'$–end is on the left with respect to the $3'$–end; solid dots indicate nucleotides that are contiguous; empty dots stand for contiguous nucleotides involved in coaxial stacking.

The recursions presented in this supplementary material are a complete description of the algorithm apart from the fact that multiloop diagrams can also have internal dangles, and coaxial diagrams can have external and internal dangles. For simplicity we omit those dangle-multiloop and dangle-coaxial diagrams from this description.

All the parameters used in the recursions—with their corresponding descriptions and values—are given in tables 2 and 3.

Symbol	Scoring parameter for	Value(Kcal/mol)
EIS^1	hairpin loops	varies
EIS^2	bulges, stems and int loops	varies
C	coaxial stacking	varies
P	external pair	0
Q	single stranded base	0
R, L	base dangling off an external pair	$dangle + Q$
P_I	pair in a nested multiloop	0.1
Q_I	not paired base inside multiloop	0.4
R_I, L_I	base dangling off a multiloop pair	$dangle + Q_I$
M	nested multiloop	4.6

Table 2: This table includes all the parameters for which there is thermodynamic information. This parameters are identical to those used in MFOLD (http://www.ibc.wustl.edu/~zuker/rna).

Symbol	Scoring parameter for	Value(Kcal/mol)
$\widetilde{EIS}^2$	IS^2 in a gap matrix	$EIS^2 * g(0.83)$
$\widetilde{C}$	coaxial stacking in pseudoknots	$C * g$
$\widetilde{P}$	pair in a pseudoknot	0.1
$\widetilde{P_I}$	pair in a non-nested multiloop	$\widetilde{P} * g$
$\widetilde{Q}$	not paired base in pseudoknot	0.2
$\widetilde{R}, \widetilde{L}$	base dangling off a pseudoknot pair	$dangle * g + \widetilde{Q}$
$\widetilde{M}$	non-nested multiloop	8.43
G_w	generating a new pseudoknot	7.0
G_{wI}	generating a pseudoknot in a multiloop	13.0
G_{wh}	overlapping pseudoknots	6.0

Table 3: In this table we introduce the new thermodynamic parameters specific for pseudoknot configurations which we had to estimate.

1 vx recursion

The recursion for the non-gap matrix vx is given by (cf. fig. 1):

$$
vx(i,j) = optimal \begin{cases}
IS^1(i,j) & \text{IS(1)} \\
IS^2(i,j:k,l) + vx(k,l) & \text{IS(2)} \\
P_I + M + wx_I(i+1,k) + wx_I(k+1,j-1) & \\
2 * P_I + C(i,j:i+1,k) + M + vx(i+1,k) + wx_I(k+1,j-1) & \text{nested} \\
2 * P_I + C(j-1,k+1:j,i) + M + wx_I(i+1,k) + vx(k+1,j-1) & \text{multiloops} \\
3 * P_I + C(k,i':k+1,j') + M + vx(i',k) + vx(k+1,j') & \\
\widetilde{P_I} + \widetilde{M} + G_{wI} + whx(i+1,r:k,l) & \\
\quad +whx(k+1,j-1:l-1,r+1) & \\
2 * \widetilde{P_I} + \widetilde{M} + G_{wI} + \widetilde{C}(i,j,:i+1,r) & \\
\quad +zhx(i+1,r:k,l) + whx(k+1,j-1:l-1,r+1) & \text{non-nested} \\
2 * \widetilde{P_I} + \widetilde{M} + G_{wI} + \widetilde{C}(j-1,k+1:j,i) & \text{multiloops} \\
\quad +whx(i+1,r:k,l) + zhx(k+1,j-1:l-1,r+1) & \\
3 * \widetilde{P_I} + \widetilde{M} + G_{wI} + \widetilde{C}(l-1,r+1:l,k) & \\
\quad +yhx(i+1,r:k,l) + yhx(k+1,j-1:l-1,r+1) &
\end{cases} \tag{1}
$$

$$[\forall i, i', k, l, r, j', j \quad i \le i' \le k \le l \le r \le j' \le j]$$

With the initialization condition

$$vx(i,i) = +\infty, \qquad \forall i \quad 1 \le i \le N. \tag{2}$$

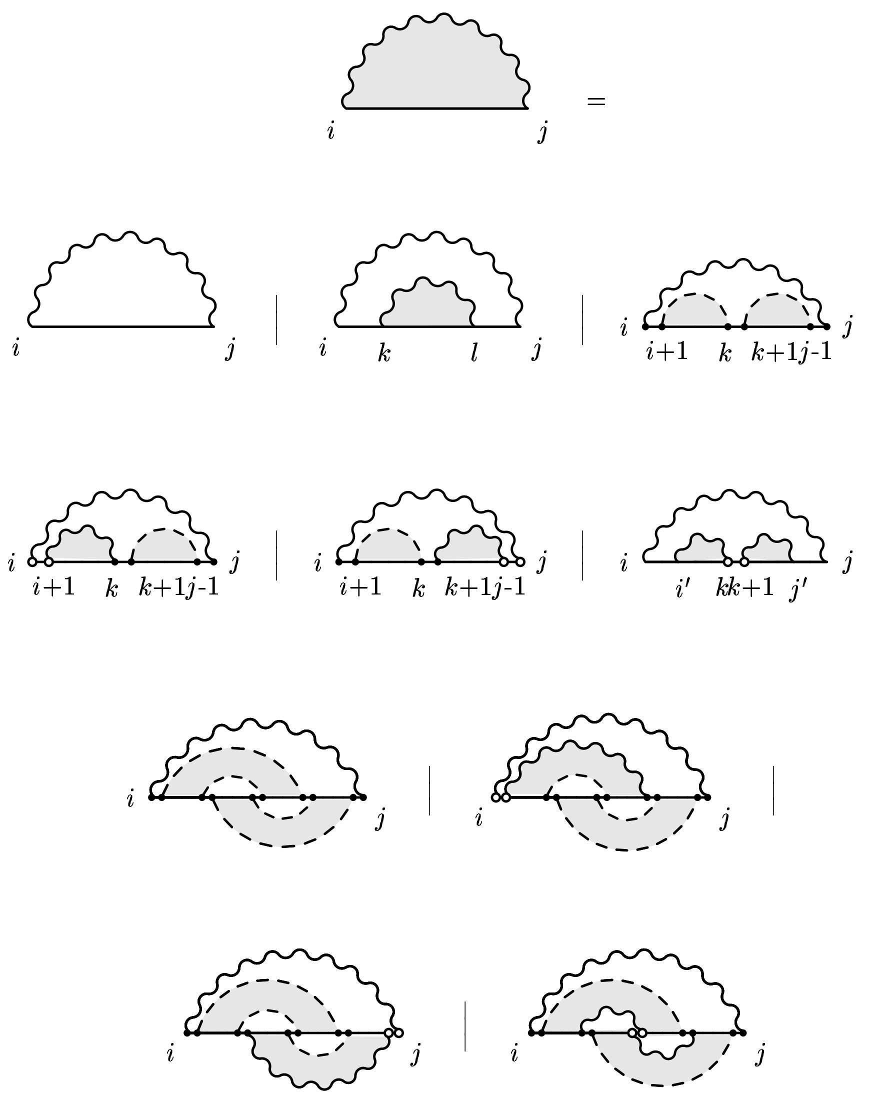


Figure 1: Recursion for the vx matrix.

2 wx recursion

The recursion for the non-gap matrix wx is given by (cf. fig. 2):

$$wx(i,j) = optimal \begin{cases} P + vx(i,j) & \text{paired} \\ L^i_{i+1,j-1} + R^j_{i+1,j-1} + P + vx(i+1,j-1) & \\ L^i_{i+1,j} + P + vx(i+1,j) & \text{dangles} \\ R^j_{i,j-1} + P + vx(i,j-1) & \\ Q + wx(i+1,j) & \text{single} \\ Q + wx(i,j-i) & \text{stranded} \\ wx(i,k) + wx(k+1,j) & \text{nested} \\ C(k,i:k+1,j) + vx(i,k) + vx(k+1,j) & \text{bifurcations} \\ G_w + whx(i,r:k,l) + whx(k+1,j:l-1,r+1) & \\ 2 * \widetilde{P} + G_w + \widetilde{C}(l-1,r+1:l,k) & \text{non-nested} \\ \quad +yhx(i,r:k,l) + yhx(k+1,j:l-1,r+1) & \text{bifurcations} \end{cases} \tag{3}$$

With the initialization conditions

$$wx(i,i) = 0, \qquad \forall i \quad 1 \le i \le N. \tag{4}$$

We should remember that the algorithm uses two different wx matrices depending on whether the subset $i...j$ is free-standing (wx) or appears inside a multiloop (in which case we use wx_I). Both recursions are identical apart from having different coefficients as described in tables 2 and 3.

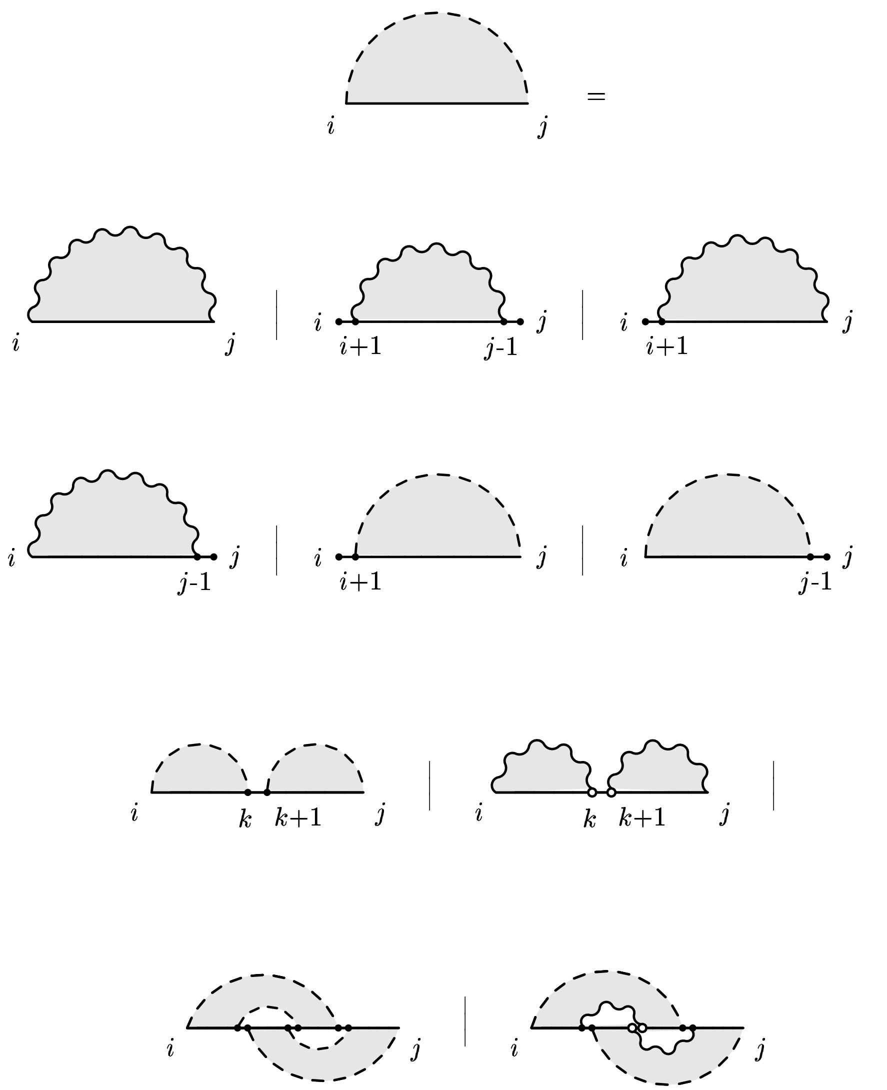


Figure 2: Recursion for the wx matrix.

3 *vhx* recursion

The recursion for the vhx matrix in the pseudoknot algorithm is given by (cf. fig. 3):

$$vhx(i,j:k,l) = optimal \begin{cases} \widetilde{EIS}^2(i,j:k,l) \\ \widetilde{EIS}^2(i,j:r,s) + vhx(r,s:k,l) \\ \widetilde{EIS}^2(r,s:k,l) + vhx(i,j:r,s) \\ 2 * \widetilde{P} + \widetilde{M} + whx(i+1,j-1:k-1,l+1) \end{cases} \tag{5}$$

$$[\forall i,r,k,l,s,j \quad i \le r \le k \le l \le s \le j]$$

The initialization conditions are,

$$vhx(i,j:k,k) = +\infty, \qquad \forall i,j,k \quad 1 \le i \le k \le j \le N. \tag{6}$$

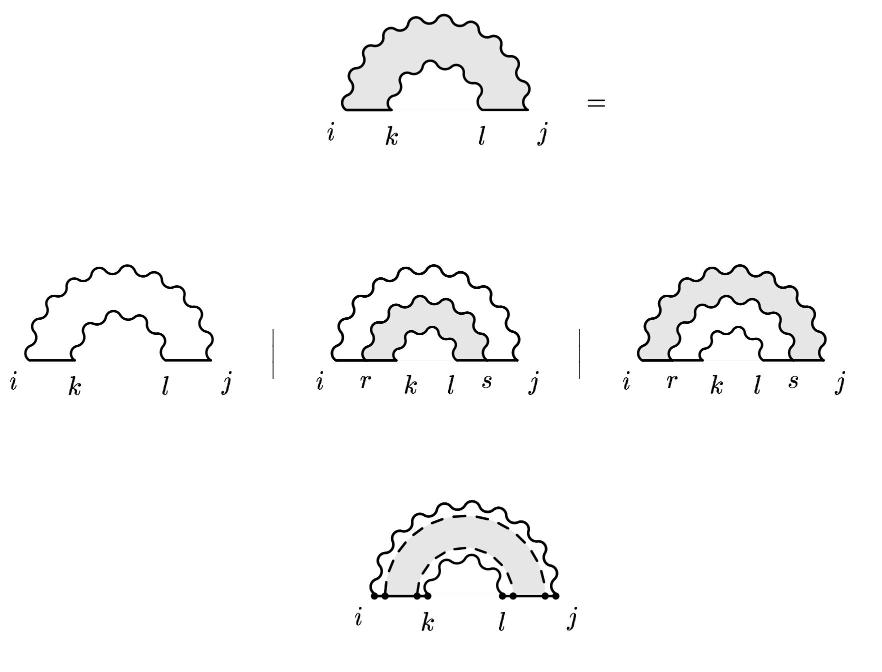


Figure 3: Recursion for the *vhx* matrix.

4 zhx and yhx recursions

The recursions for the gap matrices zhx and yhx in the pseudoknot algorithm are complementary and given by (cf. fig. 4 and fig. 5):

The recursion for the gap matrix zhx is given by (cf. fig. 4):

$$zhx(i,j:k,l) = optimal \left\{ \begin{array}{ll} \widetilde{P} + vhx(i,j:k,l) & \text{paired} \\ \widetilde{L} + \widetilde{R} + \widetilde{P} + vhx(i,j:k-1,l+1) & \\ \widetilde{R} + \widetilde{P} + vhx(i,j:k-1,l) & \text{dangles} \\ \widetilde{L} + \widetilde{P} + vhx(i,j:k,l+1) & \\ \widetilde{Q} + zhx(i,j:k-1,l) & \text{single} \\ \widetilde{Q} + zhx(i,j:k,l+1) & \text{stranded} \\ zhx(i,j:r,l) + wx_I(r+1,k) & \\ 2*\widetilde{P} + \widetilde{C}(r,l:r+1,k) + vhx(i,j:r,l) + vx(r+1,k) & \\ zhx(i,j:k,s) + wx_I(l,s-1) & \text{nested} \\ 2*\widetilde{P} + \widetilde{C}(s-1,l:s,k) + vhx(i,j:k,s) + vx(l,s-1) & \text{bifurcations} \\ \widetilde{EIS}^2(i,j:r,s) + zhx(r,s:k,l) & \\ \widetilde{P} + \widetilde{M} + whx(i+1,j-1:k,l) & \end{array} \right. \tag{7}$$

Initialization conditions,

$$zhx(i,j:k,k) = zhx(i,j:k,k+1) = vx(i,j), \quad \forall i,j,k \quad 1 \leq i \leq k \leq j \leq N. \tag{8}$$

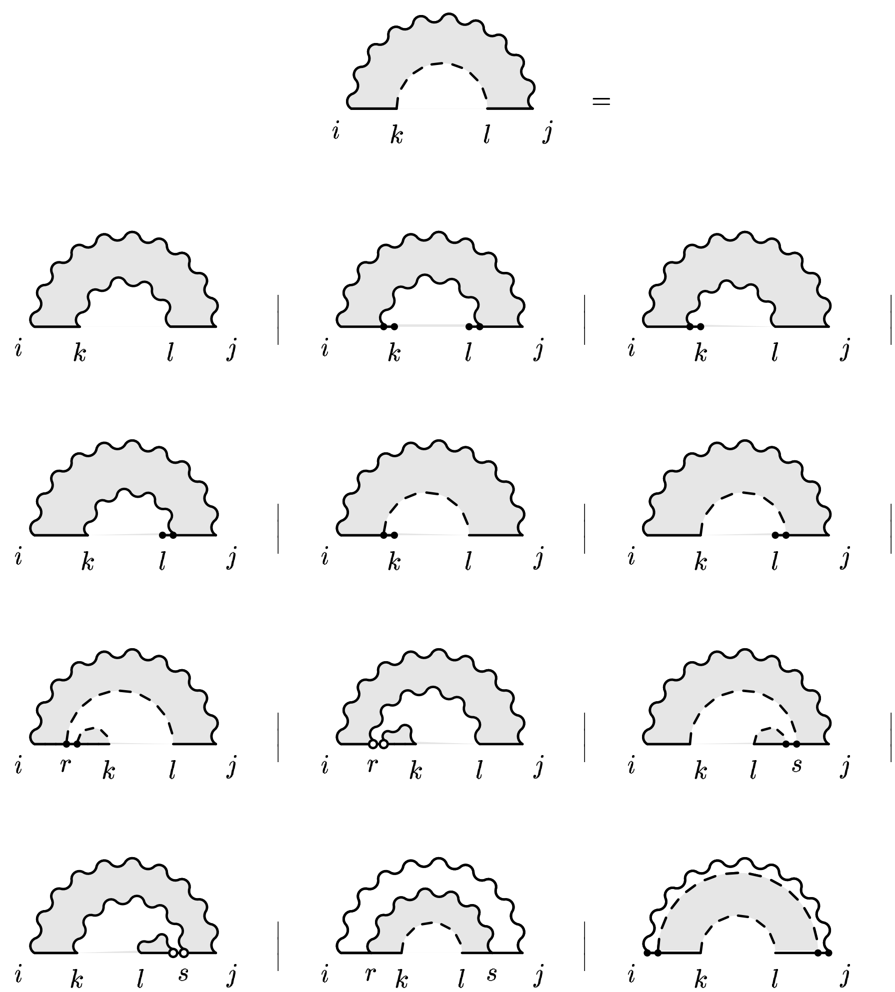


Figure 4: Recursion for the zhx matrix.

The recursion for the gap matrix yhx is given by (cf. fig. 5):

$$yhx(i,j:k,l) = optimal \left\{ \begin{array}{ll}
\widetilde{P} + vhx(i,j:k,l) & \text{paired} \\
\widetilde{L} + \widetilde{R} + \widetilde{P} + vhx(i+1,j-1:k,l) & \\
\widetilde{L} + \widetilde{P} + vhx(i+1,j:k,l) & \text{dangles} \\
\widetilde{R} + \widetilde{P} + vhx(i,j-1:k,l) & \\
\widetilde{Q} + yhx(i+1,j:k,l) & \text{single} \\
\widetilde{Q} + yhx(i,j-1:k,l) & \text{stranded} \\
wx_I(i,r) + yhx(r+1,j:k,l) & \\
2 * \widetilde{P} + \widetilde{C}(r,i:r+1,j) + vx(i,r) + vhx(r+1,j:k,l) & \\
yhx(i,s:k,l) + wx_I(s+1,j) & \text{nested} \\
2 * \widetilde{P} + \widetilde{C}(s,i:s+1,j)) + vhx(i,s:k,l) + vx(s+1,j) & \text{bifurcations} \\
yhx(i,j:r,s) + \widetilde{EIS}^2(r,s:k,l) & \\
\widetilde{P} + \widetilde{M} + whx(i,j:k-1,l+1) &
\end{array} \right. \tag{9}$$

$$[\forall i,r,k,l,s,j \quad i \leq r \leq k \leq l \leq s \leq j]$$

Initialization conditions,

$$yhx(i,j:k,k) = +\infty, \quad \forall i,j,k \quad 1 \leq i \leq k \leq j \leq N. \tag{10}$$

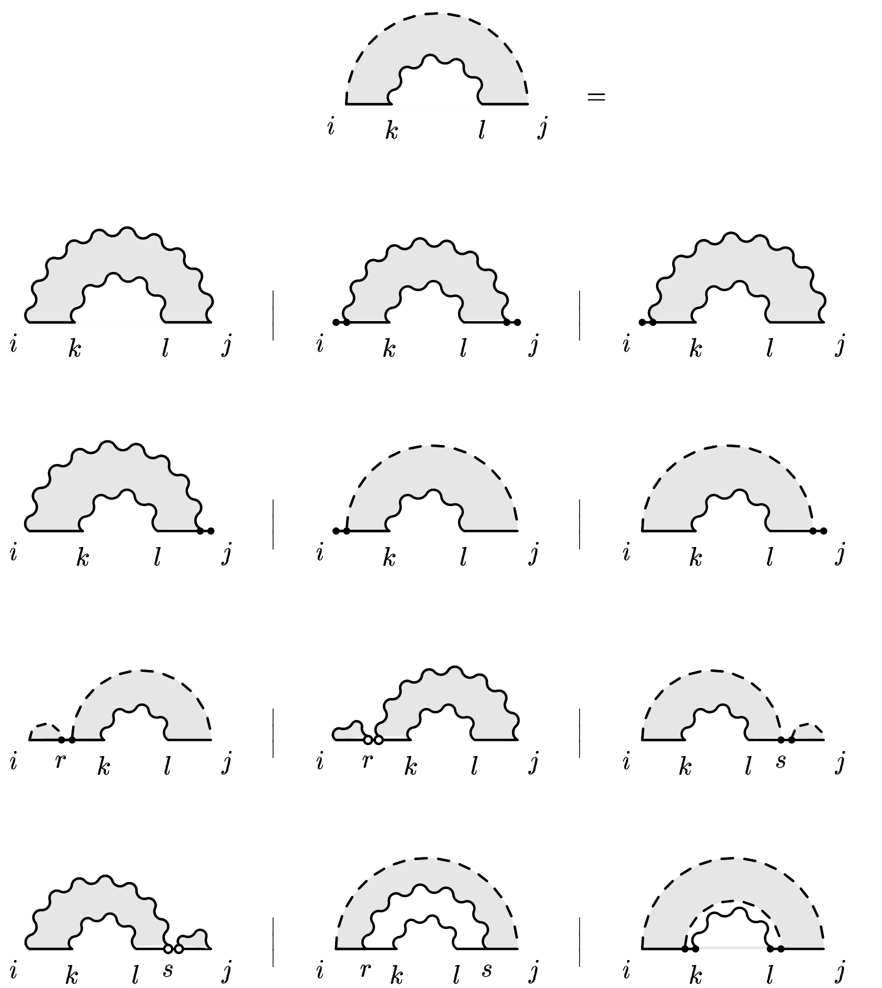


Figure 5: Recursion for the yhx matrix.

5 *whx* recursions

Finally, the recursion for the gap matrix *whx* appears in fig. 6, and is given by:

$$
whx(i,j:k,l) = optimal \left\{ \begin{array}{ll}
2*\widetilde{P} + vhx(i,j:k,l) & \\
\widetilde{P} + zhx(i,j:k,l) & \text{paired} \\
\widetilde{P} + yhx(i,j:k,l) & \\
\widetilde{L} + \widetilde{R} + 2*\widetilde{P} + vhx(i+1,j:k-1,l) & \\
\widetilde{L} + \widetilde{R} + 2*\widetilde{P} + vhx(i,j-1:k,l+1) & \\
2*\widetilde{L} + 2*\widetilde{P} + vhx(i+1,j:k,l+1) & \\
2*\widetilde{R} + 2*\widetilde{P} + vhx(i,j-1:k-1,l) & \\
\widetilde{L} + 2*\widetilde{R} + 2*\widetilde{P} + vhx(i+1,j-1:k-1,l) & \\
2*\widetilde{L} + \widetilde{R} + 2*\widetilde{P} + vhx(i+1,j:k-1,l+1) & \\
2*\widetilde{L} + \widetilde{R} + 2*\widetilde{P} + vhx(i+1,j-1:k,l+1) & \\
\widetilde{L} + 2*\widetilde{R} + 2*\widetilde{P} + vhx(i,j-1:k-1,l+1) & \text{dangles} \\
2*\widetilde{L} + 2*\widetilde{R} + 2*\widetilde{P} + +vhx(i+1,j-1:k-1,l+1) & \\
\widetilde{L} + \widetilde{R} + \widetilde{P} + zhx(i+1,j-1:k,l) & \\
\widetilde{L} + \widetilde{P} + zhx(i+1,j:k,l) & \\
\widetilde{R} + \widetilde{P} + zhx(i,j-1:k,l) & \\
\widetilde{L} + \widetilde{R} + \widetilde{P} + yhx(i,j:k-1,l+1) & \\
\widetilde{R} + \widetilde{P} + yhx(i,j:k-1,l) & \\
\widetilde{L} + \widetilde{P} + yhx(i,j:k,l+1) & \\
\widetilde{Q} + whx(i+1,j:k,l) & \\
\widetilde{Q} + whx(i,j-1:k,l) & \text{single} \\
\widetilde{Q} + whx(i,j:k-1,l) & \text{stranded} \\
\widetilde{Q} + whx(i,j:k,l+1) &
\end{array} \right.
\tag{11}
$$

(cont.)

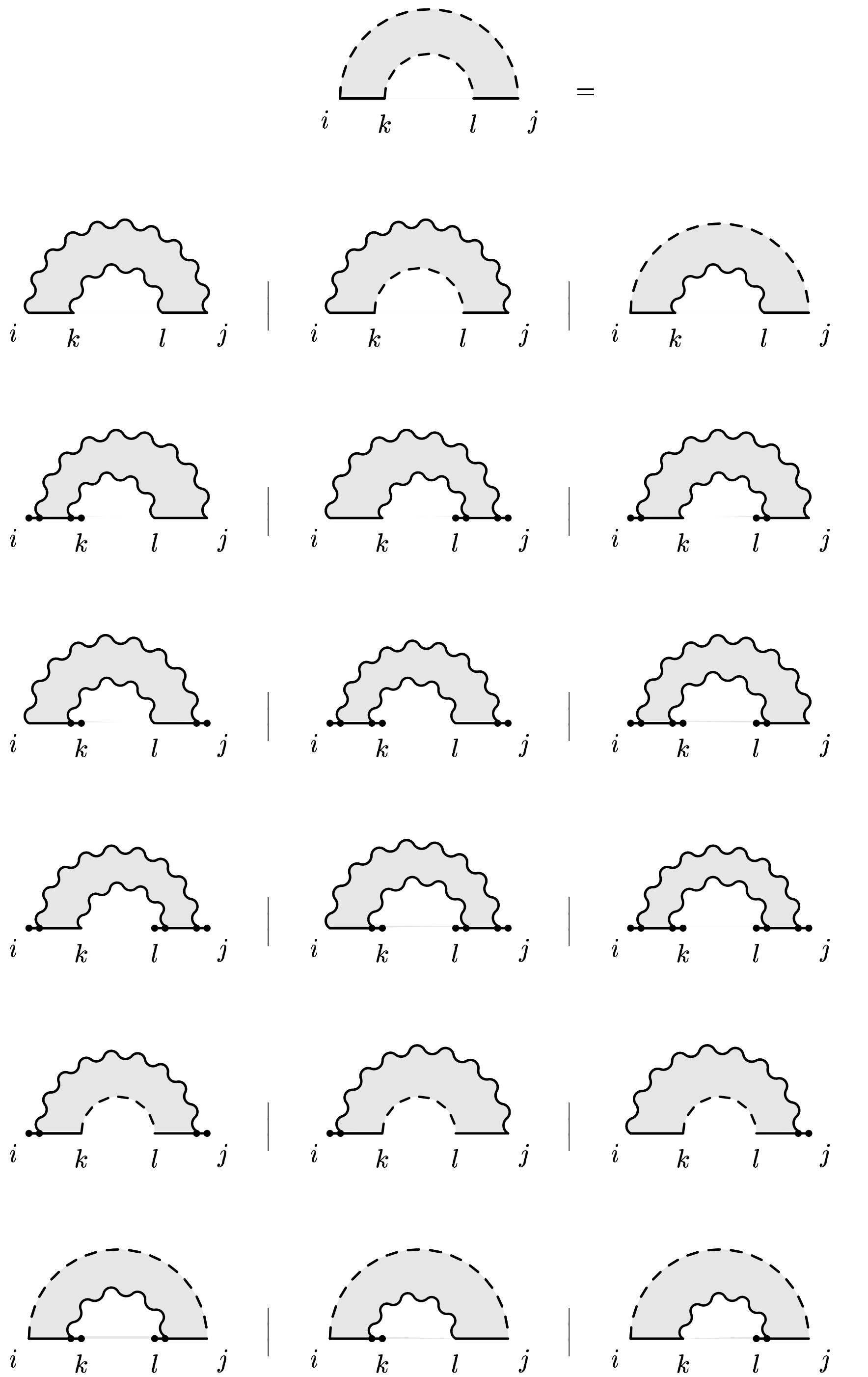
$$
whx(i,j:k,l) = optimal \begin{cases}
wx_I(i,k) + wx_I(l,j) \\
wx_I(i,r) + whx(r+1,j:k,l) \\
2 * \widetilde{P} + C(r,i:r+1,j) + vx(i,r) + zhx(r+1,j:k,l) \\
whx(i,j:r,l) + wx_I(r+1,k) \\
2 * \widetilde{P} + C(r,l:r+1,k) + yhx(i,j:r,l) + vx(r+1,k) \\
whx(i,s:k,l) + wx_I(s+1,j) \\
2 * \widetilde{P} + C(s,i:s+1,j) + zhx(i,s:k,l) + vx(s+1,j) \\
whx(i,j:k,s+1) + wx_I(l,s) \\
2 * \widetilde{P} + C(s,l:s+1,k) + yhx(i,j:k,s+1) + vx(l,s) \\
yhx(i,j:r,s) + zhx(r,s:k,l) \\
\widetilde{M} + whx(i,j:r,s) + whx(r+1,s-1:k,l) \\
\quad \text{nested bifurcations} \\
G_{wh} + whx(i,s:r,l) + whx(r+1,j:k,s+1) \\
G_{wh} + whx(i,s':k,s) + whx(l,j:s-1,s'+1) \\
2 * \widetilde{P} + G_{wh} + \widetilde{C}(s',i:s'+1,s-1) \\
\qquad + zhx(i,s':k,s) + yhx(l,j:s-1,s'+1) \\
2 * \widetilde{P} + G_{wh} + \widetilde{C}(s-1,s'+1:s,k) \\
\qquad + yhx(i,s':k,s) + yhx(l,j:s-1,s'+1) \\
G_{wh} + whx(r,j:r',l) + whx(i,k:r-1,r'+1) \\
2 * \widetilde{P} + G_{wh} + \widetilde{C}(r-1,r'+1:r,j) \\
\qquad + zhx(r,j:r',l) + yhx(i,k:r-1,r'+1) \\
2 * \widetilde{P} + G_{wh} + \widetilde{C}(r',l:r'+1,r-1) \\
\qquad + yhx(r,j:r',l) + yhx(i,k:r-1,r'+1) \\
\quad \text{non-nested bifurcations}
\end{cases} \tag{12}
$$

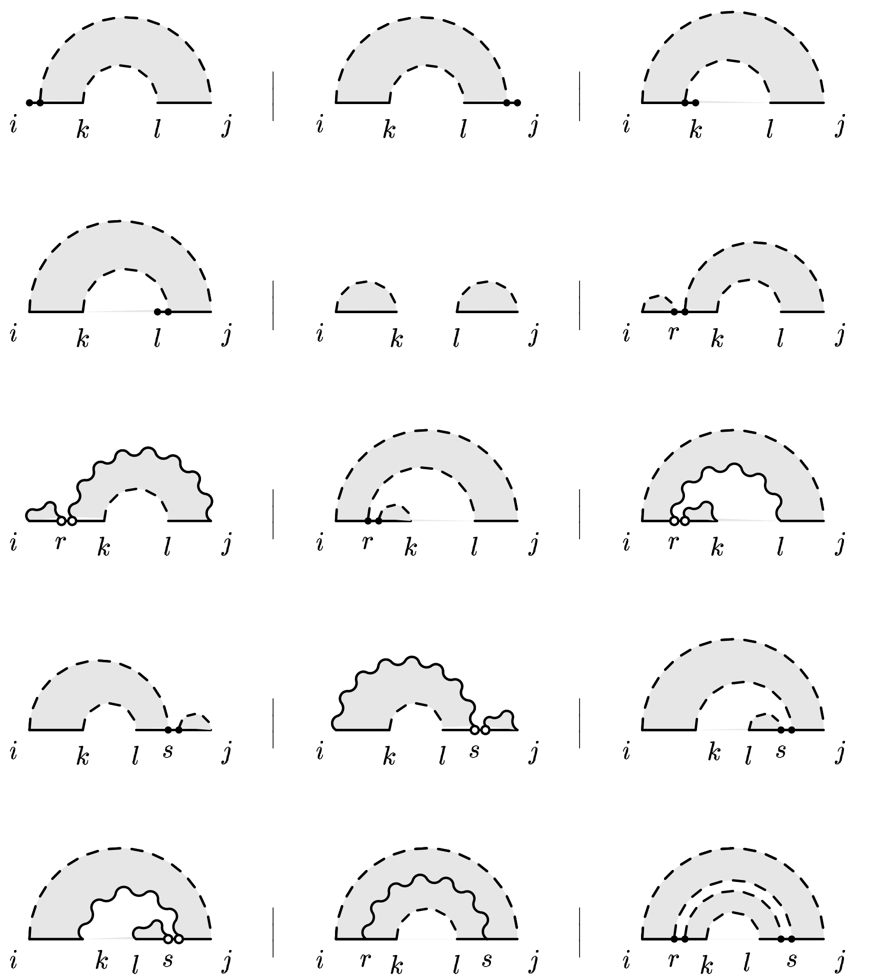
$$
[\forall i,r,r',k,l,s,s',j \quad i \le r \le r' \le k \le l \le s \le s' \le j]
$$

Initialization conditions,

$$
\begin{aligned}
whx(i,j:i,j) &= \infty, \qquad (13) \\
whx(i,j:k,k) &= whx(i,j:k,k+1) = wx(i,j),
\end{aligned}
$$

$$
\forall i,j,k \quad 1 \le i \le k \le j \le N.
$$


i k l j
=
i k l j
i k l j
i k l j
i k l j
i k l j
i k l j
i k l j
i k l j
i k l j
i k l j
i k l j
i k l j
i k l j
i k l j
i k l j
i k l j
i k l j
i k l j


i k l j
i k l j
i k l j
i k l j
i k l j
i r k l j
i r k l j
i r k l j
i r k l j
i k l s j
i k l s j
i k l s j
i k l s j
i r k l s j
i r k l s j

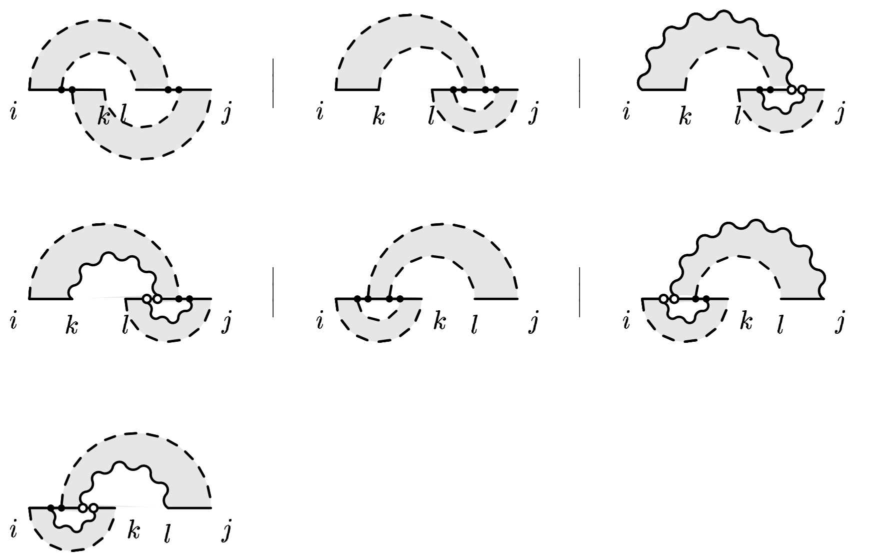


Figure 6: Recursion for the whx matrix.