Question 1:

(a) Assumption: S is a subset of R. Minimum will be M.
Assumption: No tuple in S is also in R. Maximum will be N+M.
2 marks.

(b) Assumption: No tuple in S is also in R. Minimum is 0.
Assumption: All tuples in S are also in R. Maximum is N.
2 marks

(c) As M is strictly greater than N, assuming all tuples in S are also in R means minimum is 1.
Assumption: No tuple in S is also in R. Maximum is M.
2 marks

(d) No assumptions needed. Minimum and maximum are both M x N.
2 marks

(e) Assumption: no tuple in R has a = 5. Minimum is 0.
Assumption: all tuples in R have a = 5. Maximum is M.
2 marks

(f) Minimum = ~~0~~ 1 if ~~not~~ all tuples in R have the same value for a.
Maximum is M if all tuples have distinct values for a in R.
2 marks

(g) Assumption: R and S have the same schemas, and no tuple in S appears in R. Minimum is 0.
Assumption: R and S have different schemas but no shared attributes. Maximum is M x N.
3 marks

Question 2.

(a) $R_1 = \text{Suppliers} \bowtie \text{Catalog}$

{3 marks}

$R_2 = R_1 \bowtie \text{Parts}$

$R_3 = \sigma_{colour=blue}(R_2)$

Answer $= \pi_{sname}(R_3)$

(b) $R_1 = \text{Catalog} \bowtie \text{Parts}$

{3 marks}

$R_2 = \sigma_{colour=blue}(R_1)$

$R_3 = \sigma_{colour=magenta}(R_1)$

$R_4 = \pi_{sid}(R_2)$

$R_5 = \pi_{sid}(R_3)$

Answer $= R_4 \cup R_5$

(c) $R_1 = \sigma_{colour=magenta}(\text{Parts})$

5 marks

- all magenta parts

$R_2 = \pi_{sid,pid}(R_1)$

- need only sid & pid (i.e. do not need colour anymore)

$R_3 = \pi_{pid}(R_1)$

$R_4 = \pi_{sid}(\text{Suppliers})$

$R_5 = R_4 \times R_3$

- What if all suppliers carried every magenta part?

$R_6 = R_5 - R_2$

- But we know some supplies don't. Left over tuples = not-all-magenta suppliers

$R_7 = \pi_{sid}(R_6)$

Answer $= R_4 - R_7$

- Take away supplies who don't from set of all suppliers.

Question 2 (continued)

(d) The answer here is very similar to (c), but we (in effect) eliminate the first step.

5 marks

i.e., $R_1 = Parts$

& the other expressions follow as in (c).

(e) $\rho_{R_1(sid_2,\ pid_2,\ cost)}\ Catalog$

3 marks

$R_2 = \pi_{sid2,\ pid2}(R_1)$

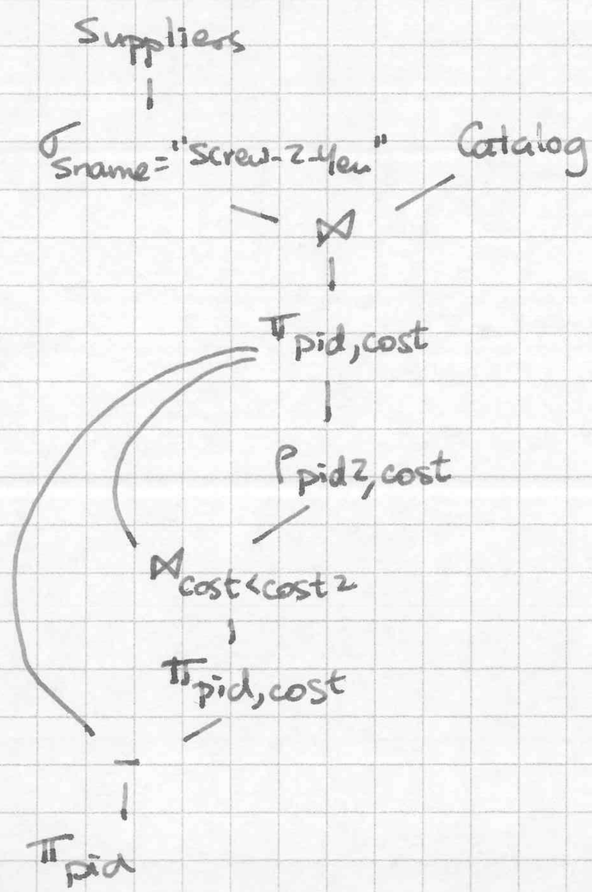
$R_3 = \pi_{sid,\ pid}(Catalog)$ $\qquad R_4 = R_2 \bowtie_{sid=sid2} R_3$

$R_5 = \sigma_{pid \neq pid2} R_4$

$answer = \pi_{pid}(R_5)$

(f)

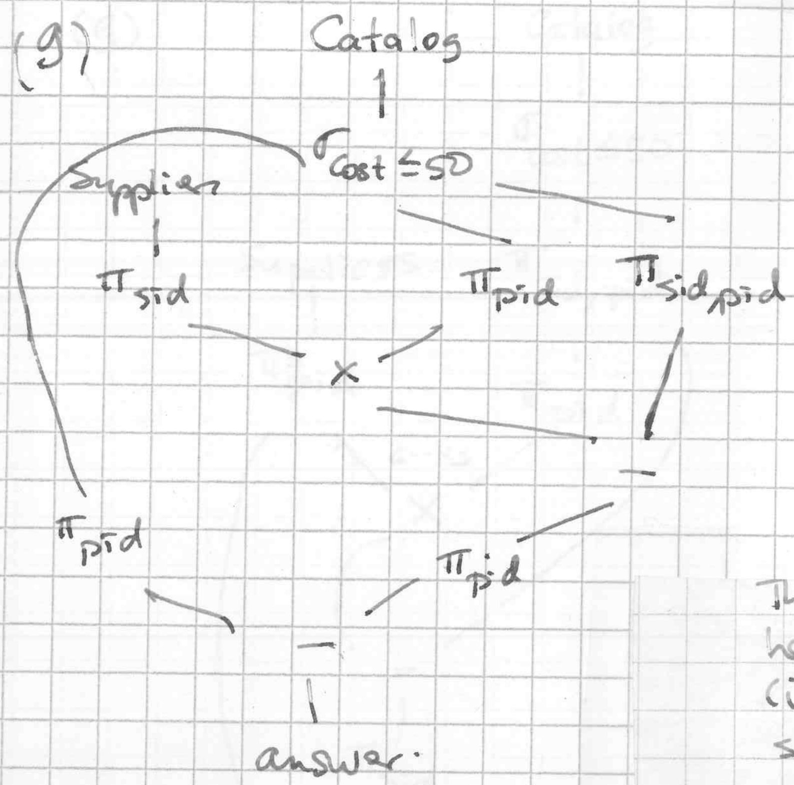
5 marks



The idea: Find all parts for which there is something more expensive. This gives us a set with all parts except the most expensive one. From there, finding the most expensive part is straight forward.

Question 2 (Continued)

(g)



(6 marks)

This is similar in idea to the expressions for (c) & (d), but now we're interested first identifying parts sold by all suppliers (rather than suppliers that sell all of a kind of part).

The cross product should probably have incoming edges swapped (i.e, to ensure our resulting schema is (sid, pid).)

Question 3 (Each part worth 5 marks)

(a) The names of all suppliers with blue parts costing less than \$100.

(b) Empty set. Projection operator at end ~~reso~~ extracts the sname attribute from a set without such an attribute.

(c) The names of all suppliers who have both cyan parts costing less than \$50 & puce parts selling for less than \$50.

(d) The ids of all suppliers selling coral- and charcoal-coloured parts for less than \$200.

(e) The names of all suppliers selling both teal-colored parts at less than \$150 and eggshell-white-coloured parts at less than \$100.

Question 4

(a) In order to derive the inferred FDs, we compute the closures of all LHS combinations using ABCD.

{15 marks}

$\{A\}^+ = \{A\}$	
$\{B\}^+ = \{B\}$	
$\{C\}^+ = \{C\}$	
$\{D\}^+ = \{D\}$	
$\{AB\}^+ = \{ABCD\}$	$AB \rightarrow D$
$\{AC\}^+ = \{AC\}$	
$\{AD\}^+ = \{ABCD\}$	$AD \rightarrow C$
$\{BC\}^+ = \{ABCD\}$	$BC \rightarrow A$
$\{BD\}^+ = \{BD\}$	
$\{CD\}^+ = \{ABCD\}$	$CD \rightarrow A$
$\{ABC\}^+ = \{ABCD\}$	$ABC \rightarrow D$
$\{ABD\}^+ = \{ABCD\}$	$ABD \rightarrow C$
$\{BCD\}^+ = \{ABCD\}$	$BCD \rightarrow A$
$\{ACD\}^+ = \{ABCD\}$	$ACD \rightarrow B$
$\{ABCD\}^+ = \{ABCD\}$	

Here we make note of the FDs inferred from the original.

(b) Examing the closures in (a) reveals the following keys:

AB, AD, BC, CD

{5 marks}

(c) Again examing the closures in (a) & keys in (b), we find five superkeys:

ABC, ABD, ACD, BCD, ABCD

{5 marks}

Question 5

15 marks

(a) There are 31 non-empty LHS combinations, and so an answer so work through these - The closures are not shown here.

From these closures — and after eliminating redundant attributes on the RHS - we find the following new FDs (The asterisks refer to a later part of this answer). (G = given)

G $AB \to C$ *	G $BC \to D$ *	$ABD \to C$ *	$BCE \to D$ *	G $CD \to A$ *
$DE \to C$ *	$BE \to C$ *	$ABE \to C$ *	$BDE \to C$ *	$AD \to B$ * G
$B \to D$ *	$BE \to D$ *	$ABE \to D$ ~~*~~	$ABCE \to D$	
$AB \to D$ *	$ABC \to D$ *	$ADE \to C$ ~~*~~	$ABDE \to C$	

In looking through the computed closures (again, not shown here) we find that $\{ABE\}^+ = \{ABCDE\}$, and further that ABE appears to be the only key.

∴ All FDs without ABE on the left will violate BCNF. Those FDs violating BCNF are marked with an asterisk.

15 marks

(b) Let us choose $AB \to C$ as the basis of our decomposition. (There are many other such choices, but this is the first of the given FDs violating BCNF, so we start with it).

$$\{AB\}^+ = \{ABCD\}$$

$$\therefore \quad R_1 = ABCD \qquad R_2 = ABE \quad [\text{i.e., } ABCDE - (ABCD + AB)]$$

Is R_1 now in BCNF? Its only key is AB (i.e., $AB \to C$, $B \to D$). However, there do exist inferred FDs that project onto the attributes of R_1 that also are BCNF violations. For example, $B \to D$ is a BCNF violation. Let us use it to decompose R_1.

$$\{B\}^+ = BD \quad \therefore \quad R_3 = BD \qquad R_4 = ABCD - (D+B) = ABC$$

Is R_3 in BCNF? Yes - the only relevant FD is $B \to D$, & B is a key for R_3.

Question 5 contd.

Is R_4 in BCNF? AB is the only key (i.e., $AB \to C$), and the FD $AB \to C$ is also the only FD projecting onto R_4. $\therefore$ R_4 is in BCNF.

Back to R_2. Is it in BCNF? There appear to be no keys for ABE amongst the FDs. Therefore the only key for R_2 is ABE itself. No FDs project onto ABE, therefore no BCNF violations are possible — R_2 is in BCNF.

Our final decomposition is: BD ABC ABE

Question 6

10 marks

(a) We could compute all 63 closures & that would show completeness. However, we can examine all the given FDs and notice that two attributes are missing from all of them: I and S. If we cannot therefore infer I or S from any of the FDs, then any key must contain I and S.

$$\{IS\}^+ \to \{IS\} \to \{ISQ\} \to \{ISQBD\} \to \{ISQBDO\}$$

$\therefore$ IS is the only key.

10 marks

(b) If we remove any one of the given FDs, then there is no way we can infer the removed FD from those that remain.

Remove $S \to D$? $\{S\}^+$ remains $\{S\}$ with other FDs.
Remove $I \to B$? $\{I\}^+$ remains $\{I\}$ with other FDs.
Remove $IS \to Q$? $\{IS\}^+ \to \{IS\} \to \{ISB\} \to \{ISBD\}$ (but no Q!)
Remove $B \to O$? $\{B\}^+ \to \{B\}$.

Could we remove one of the LHS attributes from $IS \to Q$ & still derive the rest?

Question 6 contd.

No. Neither $I \to Q$ nor $S \to Q$ in the set of FDs will help us recover $IS \to Q$ with all other FDs.

∴ The given FDs are a minimal basis.

10 marks

(c) The only key is IS, ∴ the remaining three given FDs are each a 3NF violation. To decompose into 3NF, we use the minimal basis to form the schemas:

SD IB ISQ BO

and since IS is a key, we need not construct an extra "key" schema.

To convince ourselves that the decomposition is lossless, let's use the chase.

B	O	I	S	Q	D		
b_1	o_1	i_1	s	q_1	d	from	SD
b	o_2	i	s_2	q_2	d_2	"	IB
b_3	o_3	i	s	q	d_3	"	ISQ
b	o	i_4	s_4	q_4	d_4	"	BO

After using $S \to D$, the tableau is.

B	O	I	S	Q	D
b_1	o_1	i_1	s	q_1	d
b	o_2	i	s_2	q_2	d_2
b_3	o_3	i	s	q	d
b	o	i_4	s_4	q_4	d_4

After using $I \to B$ we have

B	O	I	S	Q	D
b_1	o_1	i_1	s	q_1	d
b	o_2	i	s_2	q_2	d_2
b	o_3	i	s	q	d
b	o	i_4	s_4	q_4	d_4

Let's skip $IS \to Q$ for now.
Use $B \to O$, which yields:

B	O	I	S	Q	D
b_1	o_1	i_1	s	q_1	d
b	o	i	s_2	q_2	d_2
b	o	i	s	q	d
b	o	i_4	s_4	q_4	d_4

Applying $IS \to Q$ now produces no change to the tableau. However, the third line is what we want! The lossless join is proved.

Question 7

10 marks

(a) There are only two FDs & to construct a key we'll need to add the attribute C (i.e., $A \to D$ & $AB \to E$ gives us $AB \to DE$ but without the C). The only key is ABC.

MVDs: $\begin{cases} A \twoheadrightarrow B & \text{violates 4NF} \\ AB \twoheadrightarrow C & \text{violates 4NF} \end{cases}$

i.e., neither LHS has a key.

20 marks

(b) Let's use the first MVD as the starting point for our decomposition.

$A \twoheadrightarrow B$ implies schemas $R_1 = AB$
$R_2 = ACDE$

Is R_1 in 4NF? Yes — the only relevant MVD $A \twoheadrightarrow B$ is trivial w.r.t. R_1.

Is R_2 in 4NF? No: $A \to D$ can be promoted to $A \twoheadrightarrow D$, & this MVD (which projects onto R_2) is not 4NF.

Question 7 (contd.)

We use $A \twoheadrightarrow D$ to determine our decomposition.

$R_3 = AD$
$R_4 = ACE$

R_3 is in 4NF (cf. arguments used about R_1).

R_4 is in 4NF as no MVD projects onto the attributes of R_4. (Hence – no 4NF violations).

Final decomposition: AB, AD, ACE

Question 8

15 marks

In the text is a lovely example ~~op~~ on page 95 (Example 3.21).

R =

A	B	C
1	2	3
4	2	5

$\pi_{A,B}(R)$ =

A	B
1	2
4	2

$\pi_{B,C}(R)$ =

B	C
2	3
2	5

However, $\pi_{A,B}(R) \bowtie \pi_{B,C}(R)$ =

A	B	C
1	2	3
1	2	5
4	2	3
4	2	5

... and this is definitely <u>not</u> lossless.