Task 1.

1. For all $s \in I$, $t \mapsto f(s,t)$ is increasing.

Proof:

Fix $s \in I$, and take $t_1, t_2 \in I$. $t_1 < t_2$.

Then by problem statement $f(t_1, s) < f(t_2, s)$.

As f is symmetric, this implies:

$$f(s, t_1) < f(s, t_2)$$

$\blacksquare$

2. As μ has a finite support $S \subseteq I$, we get

$$\mathcal{L}(g_{pre}, g_{post}, \lambda) = \int_{I^2} \left(x \cdot y - \left(\lambda F(\tilde{x}, \tilde{y}) - G_{post}(x) - G_{post}(y)\right)\right)^2 d\mu =$$

$$= \int_S \left(x \cdot y - \left(\lambda F(\tilde{x}, \tilde{y}) - G_{post}(x) - G_{post}(y)\right)\right)^2 d\mu$$

as $\tilde{x} = (g_{pre}(x_i))_{1 \le i \le n}$, $G_{post}(x) = \sum_{i=1}^{n} g_{post}(x_i)$

(similarly for $\tilde{y}$, $G_{post}(y)$) only the values of g_{pre}, g_{post} on the finite set of coordinates U matter. $U = \{x_i \mid (x,y) \in S \ 1 \le i \le n\} \cup \{y_i \mid (x,y) \in S \ 1 \le i \le n\}$

So we are optimising the following finite number of parameters:

$$\{g_{pre}(u)\}_{u\in\mathcal{U}}, \{g_{post}(u)\}_{u\in\mathcal{U}}, \lambda\in\mathbb{R}.$$

4. For this problem, everything is deterministic (no noise), so we assume that $(x^{(b)}, y^{(b)})$ are pairwise distinct. μ has finite support S, and we identify B as a subset of S, $B\subseteq S$.

$$\left(s\in S \quad w_b>0, \quad \sum_{s\in S} w_s = 1, \quad \mu = \sum_{s\in S} \mathbb{1}_{(x^{(s)}, y^{(s)})}\right)$$

We can write:

$$(*) \quad \mathcal{L}(\lambda, g_{post}) = \sum_{b\in B} w_b\left(x^{(b)}y^{(b)} - \left(\lambda F(x^{(b)}, y^{(b)}) - \sum_{u\in\mathcal{U}} C_{b,u}\, g_{post}(u)\right)\right)^2$$

Proof:

as before $\mathcal{U} = \{x_s \mid (x_s, y_s), s\in S\}\cup\{y_s \mid (x_s, y_s), s\in S\}$ is set of all ~~[illegible]~~ coordinates present in the support of μ.

We have:

$$C_{post}(x^{(b)}) + C_{post}(y^{(b)}) = \sum_{i=1}^{w}\left(g_{post}(x_i^{(b)}) + g_{post}(y_i^{(b)})\right)$$

Now, as g_{post} is **only** ever used inside integrals with μ, only values $\{ g_{post}(u) \mid u \in \mathcal{U} \}$ matter.

thus: $$g_{post}(x_i^{(b)}) = \begin{cases} g_{post}(u) & \text{if } x_i^{(b)} = u \text{ for some } u \in \mathcal{U} \\ 0 & \text{otherwise (as it does not change value of integral)} \end{cases}$$

similarly for $g_{post}(y_i^{(b)})$:

Therefore:

$$C_{post}(x^{(b)}) + C_{post}(y^{(b)}) = \sum_{u \in \mathcal{U}} g_{post}(u) \left(\#\{ i \mid x_i^{(b)} = u \} + \#\{ i \mid y_i^{(b)} = u \} \right)$$

For $b \in B$, $u \in \mathcal{U}$. we introduce.

$$C_{b,u} = \#\{ i \mid x_i^{(b)} = u \} + \#\{ i \mid y_i^{(b)} = u \}$$

Then $$L(\lambda, g_{post}) = \sum_{b \in B} w_b \left(x^{(b)} y^{(b)} - \left(\lambda R(x^{(b)}, y^{(b)}) - C_{post}(x^{(b)}) - C_{post}(y^{(b)}) \right) \right)^2 =$$

$$= \sum_{b \in B} w_b \left(x^{(b)} y^{(b)} - \left(\lambda R(x^{(b)}, y^{(b)}) - \sum_{u \in \mathcal{U}} C_{b,u}\, g_{post}(u) \right) \right)^2 \blacksquare$$

Now we introduce following matrices
(as $\mathcal{B}, \mathcal{U}$ are finite, we identify them with their size)

$$b = \begin{bmatrix} x^{(1)} y^{(1)} \\ \vdots \\ x^{(B)} y^{(B)} \end{bmatrix} \quad X = \begin{bmatrix} F(x^{(1)}, y^{(1)}) & \\ \vdots & -C \\ F(x^{(B)}, y^{(B)}) & \end{bmatrix}$$

Where $C = (c_{b,u})_{b \in \mathcal{B}, u \in \mathcal{U}}$ as already defined

Also, if we define $\theta = \begin{bmatrix} \lambda \\ g_{post}(x_1) \\ \vdots \\ g_{post}(x_{card(\mathcal{U})}) \end{bmatrix}$, we get:

$$\mathcal{L}(\theta) = \| b - X\theta \|_W^2 = (b - X\theta)^T W (b - X\theta)$$

$W = \mathrm{diag}(w_b)_{b \in \mathcal{B}} > 0.$

This is least-squares problem which has minimum given by first order optimality (standard result, taking derivative)

$$X^T W X \hat{\theta} = X^T W b$$

Even though $\hat{\theta}$ may not be unique, it is global because $\mathcal{L}(\theta)$ is convex.

4

So the solution we give for $\left(x^{(b)}, y^{(b)}, P\left(x^{(b)}, y^{(b)}\right)\right)_{b \in B}$

is
$$\hat{\theta} = \begin{cases} \underbrace{(X^T W X)^{-1}}_{\text{if } X^T W X \text{ is invertible}} X^T W \mathcal{B} \\ (X^T W X)^{+} X^T W \mathcal{B} \\ \text{(using pseudo inverse, standard result, uses SVD)} \end{cases}$$

To ensure that solution is optimal:

as $W > 0$, $(X^T W X)$ invertible $\Leftrightarrow$ columns of X are independent

$\Updownarrow$

$\text{column-rank}(X) = \#\mathcal{U} + 1$

$X \in \mathbb{R}^{B \times (\#\mathcal{U} + 1)}$

But if $B \leq \#\mathcal{U}$, then $\text{column-rank}(X) \leq B \leq \#\mathcal{U}$.

Thus, $B \geq \#\mathcal{U} + 1$.

(if $\hat{\theta}$ is not unique, then we can't be sure it stays same ~~unique~~ even after we receive addional data)

$\Downarrow$

if we have $\underset{\in E_1}{\hat{\theta}_1}, \underset{\in E_2}{\hat{\theta}_2}, \ldots, \underset{\in E_i}{\hat{\theta}_i}, \underset{\in E_{i+1}}{\hat{\theta}_{i+1}}$ affine spaces

$E_1 \supseteq E_2 \supseteq \ldots$ then it could happen that

$\hat{\theta}_i \notin E_{i+1}$ (if $i \leq \#\mathcal{U}$)

3. For this question, we borrow notation from the solution of question 4. We have:

$$\mathcal{L}(q_{pre}, q_{post}) = \sum_{b \in B} w_b \left(x^{(b)} y^{(b)} - \left(F(\tilde{x}^{(b)}, \tilde{y}^{(b)}) - \sum_{u \in \mathcal{U}} c_{b,u}\, q_{post}(u) \right) \right)^2$$

With the same notation as in problem 4, but this time $F = \begin{bmatrix} F(\tilde{x}^{(1)}, \tilde{y}^{(1)}) \\ \vdots \\ F(\tilde{x}^{(B)}, \tilde{y}^{(B)}) \end{bmatrix}$

$$\theta = \left(q_{post}(u) \mid u \in \mathcal{U} \right)$$

$$\varphi = \left(q_{pre}(u) \mid u \in \mathcal{U} \right)$$

$\mathcal{L}(\varphi, \theta) = \| C\theta - (F(\varphi) - b) \|_W^2$, thus similarly to question 4:

$$\underset{\theta}{\arg\min}\, \mathcal{L}(\varphi, \theta) = (C^T W C)^+ C^T W (F(\varphi) - b) \equiv \theta^*(\varphi)$$

We are left with optimizing:

$$\varphi \mapsto \mathcal{L}(\varphi, \theta^*) = \| C\theta^* - (F(\varphi) - b) \|_W^2$$

For the remaining problem, we do know F is increasing in each coordinate, but we don't have that it is continuous ($F = \sum_{i=1}^{N} f(x_i, y_i)$ f is increasing but not necessarily continuous). Because of that, I only use finite grid method to find better solutions without guarantees that the are the best. We have

$\Phi = (g_{pre}(u) \in I$ for $u \in \mathcal{U})$. we take a finite grid $G \subset [0,1]$ $(G = \alpha_1 < \cdots < \alpha_N)$ and we find

$$\min_{\varphi \in G^{\mathcal{U}}} \varphi \longmapsto \mathcal{L}(\varphi, \theta^*(\varphi)).$$

This will have complexity $|G|^{|\mathcal{U}|}$.

5. From the solution of 4, adding the noise term, we get:

$$\mathcal{L}(\lambda, g_{post}) = \sum_{b \in B} w_b \left(x^{(b)} y^{(b)} - \left(\lambda \left(F(x^{(b)}, y^{(b)}) + \varepsilon_b \right) - \sum_{u \in U} C_{b,u}\, g_{post}(u) \right) \right)^2$$

if we use notation:

$$\alpha_b = x^{(b)} y^{(b)} - \left(\lambda F(x^{(b)}, y^{(b)}) - \sum_{u \in U} C_{b,u}\, g_{post}(u) \right)$$

Then we get

$$\mathcal{L}(\lambda, g_{post}) = \sum_{b \in B} w_b \left(\alpha_b - \lambda \varepsilon_b \right)^2$$

As the measurements are noisy, we have to optimize parameters for the best average approximation.

$$\mathbb{E}\mathcal{L}(\lambda, g_{post}) = \sum_{b \in B} w_b\, \mathbb{E}\left[(\alpha_b - \lambda \varepsilon_b)^2 \right]$$

using linearity of expection and $\mathbb{E}[\varepsilon_b] = 0$

$$\mathbb{E}\mathcal{L}(\lambda, g_{post}) = \sum_{b \in B} w_b \alpha_b^2 + \lambda^2 \sum_{b \in B} w_b \sigma_b^2$$

(we assume σ_b $b \in B$ are known, after measuring the results of the black-box enough times.)

8

Using the same matrix notation as in 2c:

$$E_L(\lambda, g_{post}) = \|b - X\theta\|_W^2 + \left(\sum_{b \in B} w_b \sigma_b^2\right)(e_1^T \theta)^2$$

The Hessian is

$$H = 2\left(X^T W X + \left(\sum_{b \in B} w_b \sigma_b^2\right) e_1 e_1^T\right)$$

For any $\vec{v}$, $\vec{v}^T H v = \|\sqrt{W} X v\|^2 + \sum_{b \in B} w_b \sigma_b^2 (e_1^T v)^2 \geq 0$

Therefore the first order conditions are sufficient and we get the following result (after derivatives)

$$\theta = \left(X^T W X + \left(\sum_{b \in B} w_b \sigma_b^2\right) e_1 e_1^T\right) X^T W b$$

This tells us that the optimization strategy does not change, we just add additional terms to address the uncertainty of data.