Allison Feaichel
ECBM 6040
HW2

Theoretical (a)

i. $$\mathcal{L}_{ls} = \sum_{i=1}^{m} (y_i - Ax_i)^T (y_i - Ax_i)$$

$$= \sum_{i=1}^{m} y_i^T y_i - 2y_i^T A x_i + (Ax_i)^T A x_i$$

$$= \sum_{i=1}^{m} y_i^T y_i - 2y_i^T A x_i + x_i^T A^T A x_i$$

$$\frac{d\mathcal{L}_{ls}}{\partial A} = \sum_{i=1}^{m} -2y_i x_i^T + 2A x_i x_i^T = 0$$

$$A \sum_{i=1}^{m} x_i x_i^T = \sum_{i=1}^{m} y_i x_i^T$$

$$A_{ls} = \left(\sum_{i=1}^{m} y_i x_i^T\right)\left(\sum_{j=1}^{m} x_j x_j^T\right)^{-1}$$

ii. $$\mathcal{L}_r = \lambda \|A\|_F^2 + \sum_{i=1}^{m} (y_i - Ax_i)^T (y_i - Ax_i)$$

$$= \lambda \operatorname{tr}(A^T A) + \sum_{i=1}^{m} y_i^T y_i - 2y_i^T A x_i + x_i^T A^T A x_i$$

$$\frac{d\mathcal{L}_r}{\partial A} = 2\lambda A + \sum_{i=1}^{m} -2y_i x_i^T + 2A x_i x_i^T = 0$$

$$A\left(\lambda + \sum_{i=1}^{m} x_i x_i^T\right) = \sum_{i=1}^{m} y_i x_i^T$$

$$A_r = \left(\sum_{i=1}^{m} y_i x_i^T\right)\left(\lambda + \sum_{j=1}^{m} x_j x_j^T\right)^{-1}$$

iii. $\epsilon_i = y_i - Ax_i$, $\epsilon_i \sim N(0, \sigma^2 I)$

$$\mathcal{L}_{ML} = \prod_{i=1}^{m} f(\epsilon_i) = \prod_{i=1}^{m} (2\pi)^{-n/2} |\sigma^2 I|^{-1/2} \exp\left\{-\tfrac{1}{2}(y_i - Ax_i)^T (\sigma^2 I)^{-1} (y_i - Ax_i)\right\}$$

$$= (2\pi\sigma^2)^{\frac{-nm}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{m} y_i^T y_i - 2y_i^T A x_i + x_i^T A^T A x_i\right\}$$

$$l_{ML} = -\frac{nm}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m} y_i^T y_i - 2y_i^T A x_i + x_i^T A^T A x_i\}$$

$$\frac{d l_{ML}}{dA} = -\frac{1}{2\sigma^2}\sum_{i=1}^{m} -2 y_i x_i^T + 2 A x_i x_i^T = 0$$

$$A \sum_{i=1}^{m} x_i x_i^T = \sum_{i=1}^{m} y_i x_i^T$$

$$A_{ML} = \left(\sum_{i=1}^{m} y_i x_i^T\right)\left(\sum_{j=1}^{m} x_j x_j^T\right)^{-1}$$

iv. $\epsilon_i = y_i - A x_i \;, \epsilon_i \sim N(0, \sigma^2 I), \quad A \sim MN(M, \lambda^{-1/2} I, \lambda^{-1/2} I)$

$M = 0I$

$$\mathcal{L}_{MAP} = \prod_{i=1}^{m} p(\epsilon_i | A) p(A) = [p(A)]^m \prod_{i=1}^{m} p(\epsilon_i | A)$$

$$= \left[\left(\frac{\lambda}{2\pi}\right)^{\frac{n^2}{2}} \exp\left\{-\frac{\lambda}{2} tr(A^T A)\right\}\right] (2\pi\sigma^2)^{\frac{-nm}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{m} (y_i - A x_i)^T (y_i - A x_i)\right\}$$

$$= \left(\frac{\lambda}{2\pi}\right)^{\frac{n^2}{2}} (2\pi\sigma^2)^{\frac{-nm}{2}} \exp\left\{-\frac{\lambda}{2} tr(A^T A) - \frac{1}{2\sigma^2}\sum_{i=1}^{m} y_i^T y_i - 2 y_i^T A x_i + x_i^T A^T A x_i\right\}$$

$$l_{MAP} = \frac{n^2}{2}\log\left(\frac{\lambda}{2\pi}\right) - \frac{mn}{2}\log(2\pi\sigma^2) - \frac{\lambda}{2} tr(A^T A) - \frac{1}{2\sigma^2}\sum_{i=1}^{m} y_i^T y_i - 2 y_i^T A x_i + x_i^T A^T A x_i$$

$$\frac{d l_{MAP}}{dA} = -\lambda A - \frac{1}{\cancel{2}\sigma^2}\sum_{i=1}^{m} -\cancel{2} y_i x_i^T + \cancel{2} A x_i x_i^T = 0$$

$$A\left(\lambda\sigma^2 I + \sum_{i=1}^{m} x_i x_i^T\right) = \sum_{i=1}^{m} y_i x_i^T$$

$$A_{MAP} = \left(\sum_{i=1}^{m} y_i x_i^T\right)\left(\lambda\sigma^2 I + \sum_{j=1}^{m} x_j x_j^T\right)^{-1}$$

v. From parts (i) and (iii) the derived expressions are the same. This is because the normalizing constant does not depend on A, the σ^2 terms in part (iii) cancel out, and raising a value to an exponent is a monotonic function, so the same value of A that minimizes part (i) also minimizes (iii)

Parts (ii) and (iv) are also almost identical answers. If (iv) assumes $\sigma^2 = 1$ then they would be equal. For the same reasons that (i) and (iii) are equivalent, parts (ii) and (iv) have solutions that are similar because the normalizing constants in (iv) do not depend on A, and the exponential is nearly the same as (ii) with the exception being (iv) has a σ^2 variable in the exponent that (ii) is missing. Another way to think of it would be, if $\epsilon_i \sim N(0, I)$ then (ii) and (iv) would have the same solution.