Advanced Algorithms: Solution of Problem 3

Comment. By no means your solutions are expected to be as long as the ones I am providing. Mine are long because I describe the discovery process.

Exercise 1

Use Remark 6 from the notes of the 2nd lecture.

Exercise 2

If g is convex, then for any $\theta \in [0, 1]$,

$$g((1-\theta)\cdot 0 + \theta \cdot 1) \leq (1-\theta)g(0) + \theta g(1)$$

Exercise 3

I start with a key special case[1]:

Special case: $A = I$

Here, $g(t) = -\ln\det((1-t)I + tB)$. As we saw in the previous homework, if the eigenvalues of B are $\lambda_1, \ldots, \lambda_n$, then the eigenvalues of $(1-t)I + tB$ are $1 - t + t\lambda_1, \ldots, 1 - t + t\lambda_n$. Since B is positive definite (PD), all λ_i are positive, and so all $1 - t + t\lambda_i$ are positive too. Thus,

$$g(t) = -\ln\left(\prod_{i=1}^{n}(1 + t(\lambda_i - 1))\right) = -\sum_{i=1}^{n}\ln(1 + t(\lambda_i - 1))$$

And also

$$g'(t) = -\sum_{i=1}^{n}\frac{\lambda_i - 1}{1 + t(\lambda_i - 1)}, \quad g''(t) = \sum_{i=1}^{n}\frac{(\lambda_i - 1)^2}{(1 + t(\lambda_i - 1))^2}$$

which gives that $g''(t) \geq 0$, for all $t \in (0, 1)$. Observe that for this proof, since t belongs in $(0, 1)$, we only need that B is positive semidefinite (PSD). Can you see why?

The general case

How to use the special case to tackle the general? The first idea is to factor out A:

$$\begin{aligned} g(t) = -\ln\det((1-t)A + tB) &= -\ln\det\left(A\left((1-t)I + tA^{-1}B\right)\right) \\ &= -\ln\left(\det(A)\det\left((1-t)I + tA^{-1}B\right)\right) \\ &= -\ln\det A - \ln\det((1-t)I + tA^{-1}B) \end{aligned}$$

[1] Another instructive special case is when both A and B are diagonal.

We can ignore the first term $-\ln\det A$, which will vanish after we take the derivative. The second term seems to be the same as in the special case, but with $A^{-1}B$ in the place of B. Are we done? No! In the proof for the special case, we heavily use the fact that B is PSD. However, $A^{-1}B$ is not necessarily symmetric! (which is a requirement for being PSD). In general, the product of symmetric matrices is not necessarily symmetric. OK, *this attempt failed, but it taught us something*: if we pull A outside, and leave in the place of B a PSD matrix, we are done. But, in Homework 2 (Problem 1.2), we saw that certain products of matrices are PSD:

Fact 1. Let $B, U \in \mathbb{R}^{n\times n}$, where $B \succcurlyeq 0$. Then, $UBU^\top \succcurlyeq 0$.

This combines perfectly with the other fact from Homework 2 (Problem 1.3):

Fact 2. If $A \in \mathbb{R}^{n\times n}$, such that $A \succcurlyeq 0$, then there exists $V \in \mathbb{R}^{n\times n}$ such that $A = VV^\top$.

Let's use Fact 2 for the matrix A in our problem. We write $A = VV^\top$. Since A is PD, it is invertible (why?), and thus V is invertible (why?). We are ready to do a factorization that will work for us:

$$\begin{aligned} g(t) = -\ln\det((1-t)A + tB) &= -\ln\det((1-t)VV^\top + tB) \\ &= -\ln\det\left(V\left((1-t)I + tV^{-1}B(V^\top)^{-1}\right)V^\top\right) \end{aligned}$$

Since $(V^\top)^{-1} = (V^{-1})^\top$ and $det(V) = det(V^\top)$, we have

$$g(t) = -\ln\left((\det(V))^2 \det\left((1-t)I + tV^{-1}B(V^{-1})^\top\right)\right) = -2\ln\det(V) - \ln\det\left((1-t)I + tM\right)$$

where $M = V^{-1}B(V^{-1})^\top \succcurlyeq 0$, from Fact 1. Now, we can simply repeat the steps from the special case, and we are done.