Problem 1. Доказать тождество Вудбери:

$$(A+UCV)^{-1}=A^{-1}-A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}.$$

Здесь $A\in\mathbb{R}^{n\times n}$, $C\in\mathbb{R}^{m\times m}$, $U\in\mathbb{R}^{n\times m}$, $V\in\mathbb{R}^{m\times n}$.

Проверим, что произведение левой и правой частей единичную матрицу:

$$\begin{aligned}&(A+UCV)(A^{-1}-A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1})=\\&=\mathrm{I}-U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}+UCVA^{-1}-UCVA^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}=\\&=\mathrm{I}+UCVA^{-1}-U(\mathrm{I}+CVA^{-1}U)(C^{-1}+VA^{-1}U)^{-1}VA^{-1}=\\&=\mathrm{I}+UCVA^{-1}-UC(C^{-1}+VA^{-1}U)(C^{-1}+VA^{-1}U)^{-1}VA^{-1}=\mathrm{I}+UCVA^{-1}-UCVA^{-1}=\mathrm{I}\end{aligned}\tag{1}$$

Problem 2. Пусть $p(\vec{x})=\mathcal{N}(\vec{x}|\vec{\mu},\Sigma)$, $p(\vec{y}|\vec{x})=\mathcal{N}(\vec{y}|A\vec{x},\Gamma)$, $A\in\mathbb{R}^{m\times n}$. Найти распределение $p(\vec{x}|\vec{y})$.

Воспользуемся формулой Байеса:

$$p(x|y)=\frac{p(y|x)p(x)}{p(y)}$$

Распределения в числителе сопряжённые, поэтому мы знаем, что постериор будет нормальным, а параметры можно пересчитать:

$$p(x|y)\propto\mathcal{N}(y|Ax,\Gamma)\cdot\mathcal{N}(x|\mu,\Sigma)=C\exp\left[-\frac{1}{2}(y-Ax)^\top\Gamma^{-1}(y-Ax)-\frac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu)\right]$$

Посмотрим на функциональный вид нормального распределения:

$$\mathcal{N}(x|\mu,\Sigma)=C\exp\left[-\frac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu)\right]=C\exp\left[-\frac{1}{2}\left(x^\top\Sigma^{-1}x-2x^\top\Sigma^{-1}\mu+\mu\top\Sigma^{-1}\mu\right)\right]$$

Теперь раскроим скобки, чтобы свести постериор к этому виду:

$$\begin{aligned}&(y-Ax)^\top\Gamma^{-1}(y-Ax)+(x-\mu)^\top\Sigma^{-1}(x-\mu)=\\&=y^\top\Gamma^{-1}y-2y^\top\Gamma^{-1}Ax+x^\top A^\top\Gamma^{-1}Ax+x^\top\Sigma^{-1}x-2x^\top\Sigma^{-1}\mu+\mu^\top\Sigma^{-1}\mu=\\&=x^\top\left(A^\top\Gamma^{-1}A+\Sigma^{-1}\right)x-2(y^\top\Gamma^{-1}A+\mu^\top\Sigma^{-1})x+\ldots\end{aligned}\tag{2}$$

Дальше можно не расписывать, ведь там остаётся константа по x.
Сразу делаем вывод, что

$$\Sigma'=\left(A^\top\Gamma^{-1}A+\Sigma^{-1}\right)^{-1}.$$

Также видим, что

$$\mu'^\top\Sigma'^{-1}=y^\top\Gamma^{-1}A+\mu^\top\Sigma^{-1}\iff\mu'^\top=(y^\top\Gamma^{-1}A+\mu^\top\Sigma^{-1})\Sigma'=(y^\top\Gamma^{-1}A+\mu^\top\Sigma^{-1})\left(A^\top\Gamma^{-1}A+\Sigma^{-1}\right)^{-1}$$

$$\implies\mu'=\left(A^\top\Gamma^{-1}A+\Sigma^{-1}\right)^{-1}(A^\top\Gamma^{-1}y+\Sigma^{-1}\mu)$$

Итого получаем

$$p(x|y)=\mathcal{N}\left(x\middle|\left(A^\top\Gamma^{-1}A+\Sigma^{-1}\right)^{-1}(A^\top\Gamma^{-1}y+\Sigma^{-1}\mu),\left(A^\top\Gamma^{-1}A+\Sigma^{-1}\right)^{-1}\right)$$

Problem 3. Пусть $p(\vec{x})=\mathcal{N}(\vec{x}|\vec{\mu},\Sigma)$, $p(\vec{y}|\vec{x})=\mathcal{N}(\vec{y}|A\vec{x},\Gamma)$. Доказать, что $p(\vec{y})=\mathcal{N}(\vec{y}|A\vec{\mu},\Gamma+A\Sigma A^T)$.

Маргинализуем $p(y)$:

$$p(y)=\int_{-\infty}^{+\infty}p(y|x)p(x)dx$$

Интеграл от произведения нормальных – это нормальное. Чтобы найти параметры, воспользуемся reparameterization trick:

$$y=Ax+\varepsilon_\Gamma,\quad\varepsilon_\Gamma\sim\mathcal{N}(\varepsilon_\Gamma|0,\Gamma)$$

Посмотрим на матожидание:

$$\mathbb{E}[y]=A\mathbb{E}[x]=A\mu$$

Теперь на дисперсию:

$$\mathbb{V}ar[y]=A\mathbb{V}ar[x]A^\top+\mathbb{V}ar[\varepsilon_\Gamma]=A\Sigma A^\top+\Gamma$$

Итого получаем

$$p(y)=\mathcal{N}(y|A\mu,A\Sigma A^\top+\Gamma)$$

Problem 4. Вычислить $\frac{\partial}{\partial X}\det(X^{-1}+A)$ (все матрицы не являются симметричными).

Известно, что

$$d\det(X) = \det(X)\langle X^{-\top}, dX\rangle$$

Тогда:

$$\det(X^{-1}+A) = \det(X^{-1}+A)\langle (X^{-1}+A)^{-\top}, d(X^{-1}+A)\rangle = \det(X^{-1}+A)\langle (X^{-1}+A)^{-\top}, dX^{-1}\rangle =$$

Также известно, что

$$dX^{-1} = -X^{-1}dXX^{-1}$$

Тогда:

$$= \det(X^{-1}+A)\langle (X^{-1}+A)^{-\top}, -X^{-1}dXX^{-1}\rangle =$$

Так как скалярное произведение тут в смысле фробениуса, то его можно представить как Trace:

$$= -\det(X^{-1}+A)Tr((X^{-1}+A)^{-\top}X^{-\top}dX^{\top}X^{-\top}) = -\det(X^{-1}+A)Tr(X^{-\top}(X^{-1}+A)^{-\top}X^{-\top}dX^{\top}) =$$

$$= -\det(X^{-1}+A)\langle X^{-\top}(X^{-1}+A)^{-\top}X^{-\top}, dX\rangle$$

$$\Longrightarrow \frac{\partial}{\partial X} = -\det(X^{-1}+A)X^{-\top}(X^{-1}+A)^{-\top}X^{-\top}$$

Problem 5. Вычислить $\frac{\partial}{\partial X}Tr(AX^{-\top}BXC)$ (все матрицы не являются симметричными, матрицы A, C не являются квадратными).

$$d\,Tr(AX^{-\top}BXC) = Tr(d(AX^{-\top}BXC)) = Tr(Ad(X^{-\top})BXC + AX^{-\top}Bd(X)C)$$

Разберёмся с каждый членом по отдельности:

1.
$$Tr(Ad(X^{-\top})BXC) = Tr(Ad(X^{-\top})BXC) = -Tr(AX^{-\top}d(X^{\top})X^{-\top}BXC) = -Tr(X^{-\top}BXCAX^{-\top}d(X^{\top}))$$

2.
$$Tr(AX^{-\top}Bd(X)C) = Tr(CAX^{-\top}Bd(X)) = Tr(B^{\top}X^{-1}A^{\top}C^{\top}d(X)^{\top})$$

Итого получаем

$$\frac{\partial}{\partial X} = B^{\top}X^{-1}A^{\top}C^{\top} - X^{-\top}BXCAX^{-\top}$$