ICML 2013–SUPPLEMENTARY MATERIAL

Sébastien Giguère, François Laviolette, Mario Marchand, Khadidja Sylla

In this supplementary material, we make use of the following notation. x_i denotes the i^{th} entry of the (column) vector $X(x)$, y_j the j^{th} entry of the (column) vector $Y(y)$, $\mathbf{V}[i;j]$ denotes the entry in position (i,j) of the matrix $\mathbf{V}$. Also, $\mathbf{V}[\,;j]$ denotes the j^{th} column of the matrix $\mathbf{V}$. Finally, $\delta_{i,j}$ denotes the delta function which gives 1 if $i=j$, and 0 otherwise.

7. Example of a distribution where the minimizer of the quadratic risk has a substantial higher error rate than the optimal classifier

We consider a simple one-dimensional binary classification problem where $\mathcal{X} = \mathbb{R}$ and $\mathcal{Y} = \{-1, +1\}$. We thus consider classifiers identified by a single scalar weight w such that the output $h_w(x)$ on an input x is given by $h_w(x) = \mathrm{sgn}(wx)$.

Consider a distribution D concentrated on four points $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)\}$. Let p_i denote the weight induced by D on x_i. Hence $\sum_{i=1}^{4} p_i = 1$. The 0/1 risk is then given by $\sum_{i=1}^{4} p_i I(h_w(x_i) \neq y_i)$ and the quadratic risk is given by $\sum_{i=1}^{4} p_i(y_i - wx_i)^2$.

Let w_r denote the value of w minimizing the quadratic risk. Since the derivative (with respect to w) of the quadratic risk must vanish at w_r, we find that it is given by the solution of $w_r \sum_{i=1}^{4} p_i x_i^2 - \sum_{i=1}^{4} p_i y_i x_i = 0$, or equivalently by

$$w_r = \frac{\sum_{i=1}^{4} p_i y_i x_i}{\sum_{i=1}^{4} p_i x_i^2}\,.$$

Now let $x_1 = \epsilon$ with $p_1 = (1-\epsilon)/2$ and $y_1 = +1$. Let $x_2 = -\epsilon$ with $p_2 = (1-\epsilon)/2$ and $y_2 = -1$. Let $x_3 = 1/\epsilon$ with $p_3 = \epsilon/2$ and $y_3 = -1$. Let $x_4 = -1/\epsilon$ with $p_4 = \epsilon/2$ and $y_4 = +1$.

Hence, with this distribution, the 0/1 risk of a classifier with a positive weight w is equal to ϵ and the 0/1 risk of a classifier with a negative weight w is equal to $1-\epsilon$. The difference tends to the maximum value of 1 when ϵ goes to zero.

However, with this distribution This gives

$$w_r = \frac{-1 + \epsilon(1-\epsilon)}{(1-\epsilon)\epsilon^2 + (1/\epsilon)}\,.$$

Hence w_r is negative for all ϵ between 0 and 1. Hence the 0/1 risk of h_{w_r} is $(1-\epsilon)$ but there exists classifiers (those with positive w) having a 0/1 risk of ϵ.

8. Proof of Equation (5)

$$\mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \|Y(y) - \mathbf{V}X(x)\|^2 \;=\; \|Y(y) - \mathbf{W}X(x)\|^2 + \sigma^2 N_{\mathcal{Y}} |X(x)\|^2, . \tag{5}$$

Proof. First, note that

$$\|Y(y) - \mathbf{V}X(x)\|^2 \;\;=\;\; \|Y(y)\|^2 \;-\; 2\langle Y(y)|\mathbf{V}X(x)\rangle \;+\; \|\mathbf{V}X(x)\|^2.$$

Let us now compute the expectation according to the posterior $Q_{\mathbf{W},\sigma}$ of these three terms.

- $\mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \|Y(y)\|^2 \;=\; \|Y(y)\|^2\,.$
- For $\mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} 2\,\langle Y(y)|\mathbf{V}X(x)\rangle$:

$$\begin{aligned}
\mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} 2\langle Y(y)|\mathbf{V}X(x)\rangle &= 2\mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \langle Y(y)|\sum_{l=1}^{N_{\mathcal{X}}} x_l \mathbf{V}[\,;l]\rangle \\
&= 2\mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \sum_{l=1}^{N_{\mathcal{X}}} \langle Y(y)|x_l \mathbf{V}[\,;l]\rangle \\
&= 2\mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \sum_{l=1}^{N_{\mathcal{X}}} \sum_{q=1}^{N_{\mathcal{Y}}} y_q \mathbf{V}[q;l] x_l \\
&= 2\sum_{l=1}^{N_{\mathcal{X}}} \sum_{q=1}^{N_{\mathcal{Y}}} y_q x_l \mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \mathbf{V}[q;l] \\
&= 2\sum_{l=1}^{N_{\mathcal{X}}} \sum_{q=1}^{N_{\mathcal{Y}}} y_q x_l \;\mathbf{W}[q;l] \\
&\;\vdots \\
&= 2\,\langle Y(y)|\mathbf{W}X(x)\rangle
\end{aligned} \tag{15}$$

- For $\mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \|\mathbf{V}X(x)\|^2$, first note that since $Q_{\mathbf{W},\sigma}$ is an *isotropic* Gaussian with mean $\mathbf{W}$ and variance σ^2, we have

$$\mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \mathbf{V}[q;l]\mathbf{V}[q;k] \;\;=\;\; \mathbf{W}[q;l]\mathbf{W}[q;k] \qquad\qquad \text{if } l \neq k\,,$$

and

$$\mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \mathbf{V}[q;l]\mathbf{V}[q;l] \;\;=\;\; \mathbf{W}[q;l] \;+\; \sigma^2\,.$$

Thus, we have

$$
\begin{aligned}
\mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \|\mathbf{V}X(x)\|^2 &= \mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \langle \mathbf{V}X(x) | \mathbf{V}X(x)\rangle & (16)\\
&= \mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \left\langle \sum_{l=1}^{N_{\mathcal{X}}} x_l \mathbf{V}[\,;l] \;\middle|\; \sum_{k=1}^{N_{\mathcal{X}}} x_k \mathbf{V}[\,;k] \right\rangle \\
&= \mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \sum_{l=1}^{N_{\mathcal{X}}}\sum_{k=1}^{N_{\mathcal{X}}} x_l x_k \left\langle \mathbf{V}[\,;l] \,|\, \mathbf{V}[\,;k] \right\rangle \\
&= \mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \sum_{l=1}^{N_{\mathcal{X}}}\sum_{k=1}^{N_{\mathcal{X}}} x_l x_k \sum_{q=1}^{N_{\mathcal{Y}}} \mathbf{V}[q;l]\mathbf{V}[q;k] \\
&= \sum_{l=1}^{N_{\mathcal{X}}}\sum_{k=1}^{N_{\mathcal{X}}} x_l x_k \sum_{q=1}^{N_{\mathcal{Y}}} \mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \mathbf{V}[q;l]\mathbf{V}[q;k] & (17)\\
&= \sum_{l=1}^{N_{\mathcal{X}}} \sum_{\substack{k\,=\,1\\ k\,\neq\, l}}^{N_{\mathcal{X}}} x_l x_k \sum_{q=1}^{N_{\mathcal{Y}}} \mathbf{W}[q;l]\mathbf{W}[q;k] \\
&\qquad + \sum_{k=1}^{N_{\mathcal{X}}} x_k x_k \sum_{q=1}^{N_{\mathcal{Y}}} (\mathbf{W}[q;l]\mathbf{W}[q;k] \,+\, \sigma^2) \\
&= \left(\sum_{l=1}^{N_{\mathcal{X}}}\sum_{k=1}^{N_{\mathcal{X}}} x_l x_k \sum_{q=1}^{N_{\mathcal{Y}}} \mathbf{W}[q;l]\mathbf{W}[q;k] \right) + \sum_{k=1}^{N_{\mathcal{X}}} x_k^2 \sum_{q=1}^{N_{\mathcal{Y}}} \sigma^2 \\
&= \|\mathbf{W}X(x)\|^2 \,+\, \sigma^2 N_{\mathcal{Y}} \sum_{k=1}^{N_{\mathcal{X}}} x_k^2 & (18)\\
&= \|\mathbf{W}X(x)\|^2 \,+\, \sigma^2 N_{\mathcal{Y}} \|X(x)\|^2 \,. & (19)
\end{aligned}
$$

From all that precedes, we then obtain:

$$
\begin{aligned}
\mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \|Y(y) - \mathbf{V}X(x)\|^2 &= \mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} \left(\|Y(y)\|^2 \,-\, 2\langle Y(y)|\mathbf{V}X(x)\rangle \,+\, \|\mathbf{V}X(x)\|^2 \right) \\
&= \|Y(y)\|^2 \,-\, 2\langle Y(y)|\mathbf{W}X(x)\rangle \,+\, \|\mathbf{W}X(x)\|^2 \,+\, \sigma^2 N_{\mathcal{Y}} \|X(x)\|^2 \\
&= \|Y(y) - \mathbf{W}X(x)\|^2 \,+\, \sigma^2 N_{\mathcal{Y}} \|X(x)\|^2 \,,
\end{aligned}
$$

and we are done. □

9. Proof of Equation (6)

Proof. Let us now prove Equation (6), which is given by

$$\mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} e^{-2\|Y(y)-\mathbf{V}X(x)\|^2} = \left[\frac{\sigma^{N_{\mathcal{X}}}}{\sqrt{1+4\sigma^2\|X(x)\|^2}}\right]^{N_{\mathcal{Y}}} e^{-\frac{2\|Y(y)-\mathbf{W}X(x)\|^2}{1+4\sigma^2\|X(x)\|^2}}\,. \tag{20}$$

We will prove Equation (20) for the case of an arbitrary vector X for which each of its component is non zero. To see that the result will also hold for the case where X has some zero-valued components, note that the result will hold by replacing X with $X+\vec{\epsilon}$, where $\vec{\epsilon}$ is a vector whose entries are all equal to ϵ for an ϵ smaller than the smallest non zero component of X. The result then comes out from the continuity with respect to X of the right-hand side of Equation (20) and by taking the limit when ϵ goes to zero.

Now, let

$$\begin{aligned} I &\stackrel{\text{def}}{=} \mathop{\mathbf{E}}_{\mathbf{V}\sim Q_{\mathbf{W},\sigma}} e^{-2\|Y(y)-\mathbf{V}X(x)\|^2} \\ &= \int \frac{d\mathbf{V}}{\left(\sigma\sqrt{2\pi}\right)^{N_{\mathcal{X}}N_{\mathcal{Y}}}}\; e^{-\frac{1}{2}\frac{\|\mathbf{V}-\mathbf{W}\|^2}{\sigma^2}}\; e^{-2\|Y(y)-\mathbf{V}X(x)\|^2}\,. \end{aligned}$$

Performing the change of variables $\mathbf{U}=\mathbf{V}-\mathbf{W}$ gives

$$I = \int \frac{d\mathbf{U}}{\left(\sigma\sqrt{2\pi}\right)^{N_{\mathcal{X}}N_{\mathcal{Y}}}}\; e^{-\frac{1}{2}\frac{\|\mathbf{U}\|^2}{\sigma^2}}\; e^{-2\|Y(y)-(\mathbf{U}+\mathbf{W})X(x)\|^2}\,.$$

Now, let $\vec{A}$ be the vector of $\mathcal{H}_{\mathcal{Y}}$ defined as

$$\vec{A} \stackrel{\text{def}}{=} Y(y) - \mathbf{W}X(x)\,, \tag{21}$$

and let us denote by A_l, the l^{th} component of the vector $\vec{A}$. Then

$$-2\|Y(y) - (\mathbf{U}+\mathbf{W})X(x)\|^2 = -2\|\vec{A}\|^2 + -2\|\mathbf{U}X(x)\|^2 + 4\langle\vec{A}\mid\mathbf{U}X(x)\rangle\,.$$

This implies that

$$I = e^{-2\|\vec{A}\|^2}\int \frac{d\mathbf{U}}{\left(\sigma\sqrt{2\pi}\right)^{N_{\mathcal{X}}N_{\mathcal{Y}}}}\; e^{-\frac{1}{2}\left(\frac{\|\mathbf{U}\|^2}{\sigma^2}+4\|\mathbf{U}X(x)\|^2-8\langle\vec{A}|\mathbf{U}X(x)\rangle\right)}\,. \tag{22}$$

9.1. An analysis of the argument of the exponential function of the integral *I*

Let

$$Q \stackrel{\text{def}}{=} \left(\frac{\|\mathbf{U}\|^2}{\sigma^2} + 4\|\mathbf{U}X(x)\|^2 - 8\langle\vec{A}\mid\mathbf{U}X(x)\rangle\right)\,. \tag{23}$$

In the following, A_l denotes the l^{th} component of the vector $\vec{A}$. Then,

$$
\begin{aligned}
Q &= \sum_{i=1}^{N_{\mathcal{X}}}\sum_{l=1}^{N_{\mathcal{Y}}}\frac{\mathbf{U}_{[l;i]}^2}{\sigma^2} + 4\|\sum_{i=1}^{N_{\mathcal{X}}}\mathbf{U}_{[\,;i]}x_i\|^2 - 8\sum_{i=1}^{N_{\mathcal{X}}}\langle\vec{A} \mid \mathbf{U}_{[\,;i]}\rangle x_i \\
&= \sum_{i=1}^{N_{\mathcal{X}}}\sum_{l=1}^{N_{\mathcal{Y}}}\frac{\mathbf{U}_{[l;i]}^2}{\sigma^2} + 4\sum_{i,j=1}^{N_{\mathcal{X}}}\sum_{l=1}^{N_{\mathcal{Y}}}\mathbf{U}_{[l;i]}x_i\mathbf{U}_{[l;j]}x_j - 8\sum_{i=1}^{N_{\mathcal{X}}}\sum_{l=1}^{N_{\mathcal{Y}}}A_l\mathbf{U}[l;i]x_i \\
&= \sum_{i=1}^{N_{\mathcal{X}}}\sum_{l=1}^{N_{\mathcal{Y}}}\frac{\mathbf{U}_{[l;i]}^2}{\sigma^2} + 4\sum_{i,j=1}^{N_{\mathcal{X}}}\sum_{l=1}^{N_{\mathcal{Y}}}\mathbf{U}_{[l;i]}x_i\mathbf{U}_{[l;j]}x_j - 8\sum_{i,j=1}^{N_{\mathcal{X}}}\sum_{l=1}^{N_{\mathcal{Y}}}\delta_{i,j}A_l\mathbf{U}_{[l;i]}x_i \\
&= \sum_{i,j=1}^{N_{\mathcal{X}}}\sum_{l=1}^{N_{\mathcal{Y}}}\left(\frac{\delta_{i,j}}{\sigma^2} + 4x_ix_j\right)\mathbf{U}_{[l;i]}\mathbf{U}_{[l;j]} - 8\sum_{i,j=1}^{N_{\mathcal{X}}}\sum_{l=1}^{N_{\mathcal{Y}}}\delta_{i,j}A_l\mathbf{U}_{[l;i]}x_i\,.
\end{aligned}
$$

Let us now define the matrix $\mathbf{N}$ of dimension $N_{\mathcal{X}}\times N_{\mathcal{Y}}$ as

$$
\mathbf{N}_{[i;j]} = \frac{\delta_{i,j}}{\sigma^2} + 4x_ix_j\,. \tag{24}
$$

Now, let

$$
\mathbf{Z}_{[l;i]} \stackrel{\text{def}}{=} \frac{\mathbf{U}_{[l;i]}}{x_i} \qquad \text{for all } l=1,..,N_{\mathcal{Y}} \text{ and } i=1,..,N_{\mathcal{X}}\,. \tag{25}
$$

Recall that, w.l.o.g., x_i is different from 0 and that $\sigma>0$.

This new change of variables gives

$$
Q = \sum_{l=1}^{N_{\mathcal{Y}}}\left(\sum_{i,j=1}^{N_{\mathcal{X}}}\mathbf{N}_{[i;j]}x_ix_j\mathbf{Z}_{[l;i]}\mathbf{Z}_{[l;j]} - 8\sum_{i=1}^{N_{\mathcal{X}}}A_lx_i^2\mathbf{Z}_{[l;i]}\right)\,. \tag{26}
$$

The following claim will transform Q in such a way that it will contain a single term including the integration variable $\mathbf{Z}$. This will be achieved by using the Fermat's difference of square argument: $(A^2-B^2)=(A-B)(A+B)$.

CLAIM 1: For any $l=1,..,N_{\mathcal{Y}}$, let

$$
B_l \stackrel{\text{def}}{=} \frac{4\sigma^2A_l}{1+4\sigma^2\|X(x)\|^2}\,.
$$

Then,

$$
Q = \sum_{l=1}^{N_{\mathcal{Y}}}\left(\sum_{i,j=1}^{N_{\mathcal{X}}}\mathbf{N}_{[i;j]}x_ix_j(\mathbf{Z}_{[l;i]}-B_l)(\mathbf{Z}_{[l;j]}-B_l)\right) - \frac{16\|A\|^2\sigma^2\|X(x)\|^2}{1+4\sigma^2\|X(x)\|^2}\,.
$$

Proof of the claim. From the definition of B_l, we have that

$$
B_l\left(x_i^2 + 4x_i^2\sigma^2\|X(x)\|^2\right) = 4A_lx_i^2\sigma^2\,.
$$

Then, since $x_i^2=\sum_{j=1}^{N_{\mathcal{X}}}\delta_{i,j}x_ix_j$ and $\|X(x)\|^2 \stackrel{\text{def}}{=} \sum_{j=1}^{N_{\mathcal{X}}}x_j^2$, we have

$$
\sum_{j=1}^{N_{\mathcal{X}}}\mathbf{N}_{[i;j]}x_ix_jB_l = 4A_lx_i^2 \tag{27}
$$

Note also that

$$\begin{aligned}
\frac{16\sigma^4 A_l^2 \|X(x)\|^2}{1+4\sigma^2\|X(x)\|^2} &= B_l^2 \|X(x)\|^2 \left(1+4\sigma^2\|X(x)\|^2\right) \\
&= B_l^2 \left(\|X(x)\|^2 + 4\sigma^2\|X(x)\|^4\right) \\
&= B_l^2 \left(\sum_{i=1}^{N_\mathcal{X}} x_i^2 \; + \; \sum_{i,j=1}^{N_\mathcal{X}} 4\sigma^2 x_i^2 x_j^2\right) \\
&= B_l^2 \left(\sum_{i,j=1}^{N_\mathcal{X}} \delta_{i,j} x_i x_j \; + \; \sum_{i,j=1}^{N_\mathcal{X}} 4\sigma^2 x_i^2 x_j^2\right) \\
&= \sum_{i,j=1}^{N_\mathcal{X}} \mathbf{N}_{[i;j]} \sigma^2 x_i x_j B_l^2 \, .
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sum_{l=1}^{N_\mathcal{Y}} \sum_{i,j=1}^{N_\mathcal{X}} \left(\mathbf{N}_{[i;j]} x_i x_j (\mathbf{Z}_{[l;i]} - B_l)(\mathbf{Z}_{[l;j]} - B_l)\right) \; - \; \frac{16\|A\|^2\sigma^2\|X(x)\|^2}{1+4\sigma^2\|X(x)\|^2} \\
&= \sum_{l=1}^{N_\mathcal{Y}} \left(\sum_{i,j=1}^{N_\mathcal{X}} \left(\mathbf{N}_{[i;j]} x_i x_j (\mathbf{Z}_{[l;i]} - B_l)(\mathbf{Z}_{[l;j]} - B_l)\right) \; - \; \frac{16 A_l^2 \sigma^2\|X(x)\|^2}{1+4\sigma^2\|X(x)\|^2}\right) \\
&= \sum_{l=1}^{N_\mathcal{Y}} \left(\sum_{i,j=1}^{N_\mathcal{X}} \left(\mathbf{N}_{[i;j]} x_i x_j (\mathbf{Z}_{[l;i]} - B_l)(\mathbf{Z}_{[l;j]} - B_l)\right) \; - \; \sum_{i,j=1}^{N_\mathcal{X}} \mathbf{N}_{[i;j]} x_i x_j B_l^2\right) \\
&= \sum_{l=1}^{N_\mathcal{Y}} \sum_{i,j=1}^{N_\mathcal{X}} \Big(\mathbf{N}_{[i;j]} x_i x_j \mathbf{Z}_{[l;i]} \mathbf{Z}_{[l;j]} \; - \; \mathbf{N}_{[i;j]} x_i x_j \mathbf{Z}_{[l;i]} B_l \; - \; \mathbf{N}_{[i;j]} x_i x_j B_l \mathbf{Z}_{[l;j]} \\
&\qquad + \mathbf{N}_{[i;j]} x_i x_j B_l^2 \; - \; \mathbf{N}_{[i;j]} x_i x_j B_l^2\Big) \\
&= \sum_{l=1}^{N_\mathcal{Y}} \sum_{i,j=1}^{N_\mathcal{X}} \left(\mathbf{N}_{[i;j]} x_i x_j B_l \mathbf{Z}_{[l;i]} \mathbf{Z}_{[l;j]} \; - \; \mathbf{N}_{[i;j]} x_i x_j \mathbf{Z}_{[l;i]} B_l \; - \; \mathbf{N}_{[i;j]} x_i x_j \mathbf{Z}_{[l;j]} B_l\right) \\
&= \sum_{l=1}^{N_\mathcal{Y}} \sum_{i,j=1}^{N_\mathcal{X}} \left(\mathbf{N}_{[i;j]} x_i x_j B_l \mathbf{Z}_{[l;i]} \mathbf{Z}_{[l;j]} \; - \; 2\mathbf{N}_{[i;j]} x_i x_j \mathbf{Z}_{[l;i]} B_l\right) \\
&= \sum_{l=1}^{N_\mathcal{Y}} \left(\sum_{i,j=1}^{N_\mathcal{X}} \mathbf{N}_{[i;j]} x_i x_j \mathbf{Z}_{[l;i]} \mathbf{Z}_{[l;j]} \; - \; 2\sum_{i=1}^{N_\mathcal{X}} \left(\sum_{j=1}^{N_\mathcal{X}} \mathbf{N}_{[i;j]} x_i x_j \mathbf{Z}_{[l;i]} B_l\right)\right) \\
&= \sum_{l=1}^{N_\mathcal{Y}} \left(\sum_{i,j=1}^{N_\mathcal{X}} \mathbf{N}_{[i;j]} x_i x_j \mathbf{Z}_{[l;i]} \mathbf{Z}_{[l;j]} \; - \; 2\sum_{i=1}^{N_\mathcal{X}} 4 A_l \mathbf{Z}_{[l;i]} x_i^2\right) \\
&= Q \, .
\end{aligned}$$

The penultimate equality comes from Equation (27). Thus, Claim 1 is proved.

9.2. Let us transform our integral I into a Gaussian integral

Definition 7.

- *Let the operator* $\star\colon \{1,..,N_{\mathcal{Y}}\}\times\{1,..,N_{\mathcal{X}}\} \longrightarrow \{1,..,N_{\mathcal{Y}}N_{\mathcal{X}}\}$ *be defined as*

$$l\star i \;\stackrel{def}{=}\; (l-1)\cdot N_{\mathcal{X}} \,+\, i\,.$$

Note that for any $\tilde{l} \in \{1,..,N_{\mathcal{Y}}N_{\mathcal{X}}\}$ *there existe a unique 2-tuple* $(l,i) \in \{1,..,N_{\mathcal{Y}}\}\times\{1,..,N_{\mathcal{X}}\}$ *such that* $\tilde{l} = l\star i\,.$

- *Let* $\vec{z}$ *be the vector of dimension* $N_{\mathcal{Y}}N_{\mathcal{X}}$ *defined as*

$$z_{l\star i} \;\stackrel{def}{=}\; \mathbf{Z}_{[l;i]}$$

for any $l \in \{1,..,N_{\mathcal{Y}}\}$ *, and any* $i \in \{1,..,N_{\mathcal{X}}\}\,.$

- *Let* $\vec{\mu}$ *be the vector of dimension* $N_{\mathcal{Y}}N_{\mathcal{X}}$ *defined as*

$$\mu_{l\star i} \;\stackrel{def}{=}\; B_l$$

for any $l \in \{1,..,N_{\mathcal{Y}}\}$ *, and any* $i \in \{1,..,N_{\mathcal{X}}\}\,.$

- *Let* $\mathbf{M}$ *be the matrix of dimension* $(N_{\mathcal{Y}}N_{\mathcal{X}})\times(N_{\mathcal{Y}}N_{\mathcal{X}})$ *defined as*

$$\mathbf{M}_{[l\star i\,;\,m\star j]} \;\stackrel{def}{=}\; \delta_{l,m}\mathbf{N}_{[i;j]}x_ix_j \quad \left(=\delta_{l,m}\Big(\frac{\delta_{i,j}}{\sigma^2}\,+\,4x_ix_j\Big)x_ix_j\right)\,, \tag{28}$$

for any $l,m \in \{1,..,N_{\mathcal{Y}}\}$ *, and any* $i,j \in \{1,..,N_{\mathcal{X}}\}\,.$

Note that in what follows, the reader should interpret $\tilde{l}$ as $l\star i$ and $\tilde{m}$ as $m\star j$.

From the definitions above, we have

$$\begin{aligned}
Q &= \sum_{l=1}^{N_{\mathcal{Y}}}\left(\sum_{i,j=1}^{N_{\mathcal{X}}}\mathbf{N}_{[i;j]}x_ix_j(\mathbf{Z}_{[l;i]}-B_l)(\mathbf{Z}_{[l;j]}-B_l)\right) - \frac{16\|A\|^2\sigma^2\|X(x)\|^2}{1+4\sigma^2\|X(x)\|^2}\\
&= \sum_{m=1}^{N_{\mathcal{Y}}}\left(\sum_{l=1}^{N_{\mathcal{Y}}}\sum_{i,j=1}^{N_{\mathcal{X}}}\big(\delta_{l,m}\mathbf{N}_{[i;j]}x_ix_j(\mathbf{Z}_{[l;i]}-B_l)(\mathbf{Z}_{[l;j]}-B_l)\big)\right) - \frac{16\|A\|^2\sigma^2\|X(x)\|^2}{1+4\sigma^2\|X(x)\|^2}\\
&= \sum_{l=1}^{N_{\mathcal{Y}}}\sum_{i=1}^{N_{\mathcal{X}}}\sum_{m=1}^{N_{\mathcal{Y}}}\sum_{j=1}^{N_{\mathcal{X}}}\big(\delta_{l,m}\mathbf{N}_{[i;j]}x_ix_j(\mathbf{Z}_{[l;i]}-B_l)(\mathbf{Z}_{[l;j]}-B_l)\big) - \frac{16\|A\|^2\sigma^2\|X(x)\|^2}{1+4\sigma^2\|X(x)\|^2}\\
&= \sum_{\tilde{l}=1}^{N_{\mathcal{Y}}N_{\mathcal{X}}}\sum_{\tilde{m}=1}^{N_{\mathcal{Y}}N_{\mathcal{X}}}\Big((z_{\tilde{l}}-\mu_{\tilde{l}})\;\mathbf{M}_{[\tilde{l};\tilde{m}]}\;(z_{\tilde{m}}-\mu_{\tilde{m}})\Big) - \frac{16\|A\|^2\sigma^2\|X(x)\|^2}{1+4\sigma^2\|X(x)\|^2}\,.
\end{aligned}$$

Substituing this expression for Q into the integral I given by Equation (22) gives

$$I \;=\; e^{-2\|\vec{A}\|^2}\int\frac{d\mathbf{U}}{\big(\sigma\sqrt{2\pi}\big)^{N_{\mathcal{X}}N_{\mathcal{Y}}}}\; e^{-\frac{1}{2}\left(\frac{\|\mathbf{U}\|^2}{\sigma^2}+4\|\mathbf{U}X(x)\|^2-8\langle\vec{A}|\mathbf{U}X(x)\rangle\right)} \tag{29}$$

$$= e^{-2\|\vec{A}\|^2} \prod_{i=1}^{N_\mathcal{X}} |x_i|^{N_\mathcal{Y}} \left(\int \frac{d\vec{z}}{\left(\sigma\sqrt{2\pi}\right)^{N_\mathcal{X} N_\mathcal{Y}}} \; e^{-\frac{1}{2} \sum_{\tilde{l}=1}^{N_\mathcal{Y} N_\mathcal{X}} \sum_{\tilde{m}=1}^{N_\mathcal{Y} N_\mathcal{X}} \left((z_{\tilde{l}} - \mu_{\tilde{l}}) \; \mathbf{M}_{[\tilde{l};\tilde{m}]} \; (z_{\tilde{m}} - \mu_{\tilde{m}}) \right)} \right) \cdot e^{\frac{8\|\vec{A}\|^2 \sigma^2 \|X(x)\|^2}{1+4\sigma^2\|X(x)\|^2}} \tag{30}$$

$$= e^{-2\|\vec{A}\|^2} \prod_{i=1}^{N_\mathcal{X}} |x_i|^{N_\mathcal{Y}} \; e^{\frac{8\|\vec{A}\|^2 \sigma^2 \|X(x)\|^2}{1+4\sigma^2\|X(x)\|^2}} \int \frac{d\vec{z}}{\left(\sigma\sqrt{2\pi}\right)^{N_\mathcal{X} N_\mathcal{Y}}} \; e^{-\frac{1}{2} \left((\vec{z}-\vec{\mu})^\intercal \, \mathbf{M} \, (\vec{z}-\vec{\mu})\right)} \tag{31}$$

$$= e^{-2\|\vec{A}\|^2} \prod_{i=1}^{N_\mathcal{X}} |x_i|^{N_\mathcal{Y}} \; e^{\frac{8\|\vec{A}\|^2 \sigma^2 \|X(x)\|^2}{1+4\sigma^2\|X(x)\|^2}} \; \frac{1}{\sqrt{\det(\mathbf{M})}} \cdot \int \frac{d\vec{z}}{\left(\sigma\sqrt{2\pi}\right)^{N_\mathcal{X} N_\mathcal{Y}}} \; \frac{1}{\sqrt{\det(\mathbf{M}^{-1})}} \; e^{-\frac{1}{2} \left((\vec{z}-\vec{\mu})^\intercal \, (\mathbf{M}^{-1})^{-1} \, (\vec{z}-\vec{\mu})\right)} \tag{32}$$

$$= e^{-2\|\vec{A}\|^2} \prod_{i=1}^{N_\mathcal{X}} |x_i|^{N_\mathcal{Y}} \; e^{\frac{8\|\vec{A}\|^2 \sigma^2 \|X(x)\|^2}{1+4\sigma^2\|X(x)\|^2}} \; \frac{1}{\sqrt{\det(\mathbf{M})}} \cdot 1 \, . \tag{33}$$

Line (30) is a consequence of the fact that $\mathbf{U}_{[l;i]} = x_i \mathbf{Z}_{[l;i]}$ (see Equation (25)) and of the fact that $\vec{z}_{\tilde{l}} = \mathbf{Z}_{[l;i]}$. Line (33) comes from the fact that the integral of the preceeding line is an integral of a Gaussian density and is therefore equal to 1. Lines (32) and (33) force $\mathbf{M}$ to be positive definite, so we have to prove that fact. This is one of the statements of the following claim.

CLAIM 2: Matrix $\mathbf{M}$ *is positive definite and*

$$\det(\mathbf{M}) \;=\; \prod_{i=1}^{N_\mathcal{X}} (x_i^2)^{N_\mathcal{Y}} \left(\frac{1}{\sigma^2}\right)^{N_\mathcal{X} N_\mathcal{Y}} \left(1 + 4\sigma^2 \|X(x)\|^2\right)^{N_\mathcal{Y}}$$

Before proving Claim 2, let us show that it implies the result.

$$\begin{aligned}
I &= e^{-2\|\vec{A}\|^2} \prod_{i=1}^{N_\mathcal{X}} |x_i|^{N_\mathcal{Y}} \; e^{\frac{8\|\vec{A}\|^2 \sigma^2 \|X(x)\|^2}{1+4\sigma^2\|X(x)\|^2}} \; \frac{1}{\sqrt{\det(\mathbf{M})}} \\
&= e^{-2\|\vec{A}\|^2} \prod_{i=1}^{N_\mathcal{X}} |x_i|^{N_\mathcal{Y}} \; e^{\frac{8\|\vec{A}\|^2 \sigma^2 \|X(x)\|^2}{1+4\sigma^2\|X(x)\|^2}} \; \frac{1}{\sqrt{\prod_{i=1}^{N_\mathcal{X}} (x_i^2)^{N_\mathcal{Y}} \; \left(\frac{1}{\sigma^2}\right)^{N_\mathcal{X} N_\mathcal{Y}} \left(1 + 4\sigma^2 \|X(x)\|^2\right)^{N_\mathcal{Y}}}} \\
&= e^{-2\|\vec{A}\|^2} \; e^{\frac{8\|\vec{A}\|^2 \sigma^2 \|X(x)\|^2}{1+4\sigma^2\|X(x)\|^2}} \; \frac{1}{\sqrt{\left(\frac{1}{\sigma^2}\right)^{N_\mathcal{X} N_\mathcal{Y}} \left(1 + 4\sigma^2 \|X(x)\|^2\right)^{N_\mathcal{Y}}}} \\
&= e^{-2\|\vec{A}\|^2} \; e^{\frac{8\|\vec{A}\|^2 \sigma^2 \|X(x)\|^2}{1+4\sigma^2\|X(x)\|^2}} \; \frac{\sigma^{N_\mathcal{X} N_\mathcal{Y}}}{\sqrt{\left(1 + 4\sigma^2 \|X(x)\|^2\right)^{N_\mathcal{Y}}}} \\
&= e^{\frac{-2\|\vec{A}\|^2}{1+4\sigma^2\|X(x)\|^2}} \; \frac{\sigma^{N_\mathcal{X} N_\mathcal{Y}}}{\sqrt{\left(1 + 4\sigma^2 \|X(x)\|^2\right)^{N_\mathcal{Y}}}} \\
&= e^{\frac{-2\|Y(y) - \mathbf{W}X(x)\|^2}{1+4\sigma^2\|X(x)\|^2}} \; \frac{\sigma^{N_\mathcal{X} N_\mathcal{Y}}}{\sqrt{\left(1 + 4\sigma^2 \|X(x)\|^2\right)^{N_\mathcal{Y}}}} \, .
\end{aligned}$$

To finish the proof, let us now prove Claim 2.

Proof of the claim. Let $\mathbf{X}$ be the diagonal matrix whose entries are the x_is and note that the matrix $(\mathbf{N}_{[i;j]}x_ix_j)_{i\,;\,j}$ can be expressed as follows:

$$(\mathbf{N}_{[i;j]}x_ix_j)_{i\,;\,j} \quad = \quad \mathbf{XNX}\,. \tag{34}$$

Now, from the definition of $\mathbf{M}$, and basic determinant's properties, we have

$$\begin{aligned}
\det(\mathbf{M}) \quad &= \quad \det\Big((\delta_{l,m}\,\mathbf{N}_{[i;j]}x_ix_j)_{l\star i\,;\,m\star j}\Big) && (35)\\
&= \quad \Big(\det\big((\mathbf{N}_{[i;j]}x_ix_j)_{i\,;\,j}\big)\Big)^{N_{\mathcal{Y}}} && (36)\\
&= \quad \Big(\det\big(\mathbf{XNX}\big)\Big)^{N_{\mathcal{Y}}} && (37)\\
&= \quad \left(\left(\prod_{i=1}^{N_{\mathcal{X}}} x_i\right)\left(\prod_{j=1}^{N_{\mathcal{X}}} x_j\right)\ \det\Big(\mathbf{N}\Big)\right)^{N_{\mathcal{Y}}} && (38)\\
&= \quad \left(\left(\prod_{i=1}^{N_{\mathcal{X}}} x_i^2\right)\ \det(\mathbf{N})\right)^{N_{\mathcal{Y}}}
\end{aligned}$$

Line (35) comes straightforwardly from the definition (see Equation (28)). Line (36) comes from the fact that $\mathbf{M}$ is a matrix whose entries are all 0, except for $N_{\mathcal{Y}}$ identical blocks of size $N_{\mathcal{X}} \times N_{\mathcal{X}}$ that are positioned in the diagonal of M, each one of those blocks being the matrix $(\mathbf{N}_{[i;j]}x_ix_j)_{i\,;\,j}$. Line (38) follows from a basic determinant's property, and from the fact that $\det(\mathbf{X}) = \Big(\prod_{i=1}^{N_{\mathcal{X}}} x_i\Big)$.

Note also that the block structure of the matrix $\mathbf{M}$ implies that it has exactly the same eigenvalues as Matrix $(\mathbf{N}_{[i;j]}x_ix_j)_{i\,;\,j}$ (but with a multiplicity augmented by a factor of $N_{\mathcal{Y}}$).

Also, it follows from Equation (34) that, for each eigenvalue λ of $(\mathbf{N}_{[i;j]}x_ix_j)_{i\,;\,j}$, there exists i such that $\frac{\lambda}{x_i^2}$ is an eigenvalue of $\mathbf{N}$. Indeed, because of Equation (34), we have that

$$\det\Big((\mathbf{N}_{[i;j]}x_ix_j)_{i\,;\,j} - \lambda\mathbf{XX}\Big) = \mathbf{0} \quad\Leftrightarrow\quad \det\Big(\mathbf{N} - \lambda\,I\Big) = \mathbf{0}\,.$$

This, in turn, implies that if $\mathbf{N}$ is positive definite, so is $\mathbf{M}$.

Hence, to prove Claim 2, we only have to show that $\mathbf{N}$ is positive definite and

$$\det(\mathbf{N}) = \left(\frac{1}{\sigma^2}\right)^{N_{\mathcal{X}}}\left(1 + 4\sigma^2\|X(x)\|^2\right)\,.$$

Let us consider matrix $\mathbf{O}$, defined as $\mathbf{O}_{[i;j]} = 4x_ix_j$. Then, it is easy to see that $\lambda = 0$ is an eigenvalue of $\mathbf{O}$ of multiplicity $N_{\mathcal{X}} - 1$ because the rank of that matrix is 1. Note that line L_i of that matrix is always equal to $\frac{x_i}{x_1}L_1$. Moreover we can easily see that $(x_1, \ldots, x_m)^\intercal$ is an eigenvector of $\mathbf{O}$ with eigenvalue $4\|X(x)\|^2$.

Now, note that

$$\mathbf{N} = \mathbf{O} + \frac{1}{\sigma^2}\cdot I\,.$$

Thus, there is a one-to-one correspondence between the eigenvalues of $\mathbf{O}$ and those of $\mathbf{N}$: λ is an eigenvalue of the former if and only if $\lambda + \frac{1}{\sigma^2}$ is an eigenvalue of the latter. Thus N is positive definite, and

$$\begin{aligned}
\det(\mathbf{N}) \quad &= \quad \left(\frac{1}{\sigma^2}\right)^{N_{\mathcal{X}}-1}\left(\frac{1}{\sigma^2} + 4\|X(x)\|^2\right)\\
&= \quad \left(\frac{1}{\sigma^2}\right)^{N_{\mathcal{X}}}\left(1 + 4\sigma^2\|X(x)\|^2\right)\,.
\end{aligned}$$

□

10. Proof of $\frac{\partial}{\partial \mathbf{A}} R(\mathbf{A}, S)$ from Theorem (6)

Proof. From equation (9) we have

$$\mathbf{W} \;=\; \sum_{i=1}^{m}\sum_{j=1}^{m} Y(y_i)A_{[i;j]}X^{\dagger}(x_j) = \mathbf{M}_{\mathcal{Y}}\mathbf{A}\mathbf{M}_{\mathcal{X}}^{\dagger} \tag{39}$$

Where $M_{\mathcal{Y}}$ is a $N_{\mathcal{Y}} \times m$ matrix with $Y(y_i)$ in it's i-th column. Similarly $M_{\mathcal{X}}$ is a $N_{\mathcal{X}} \times m$ matrix with $X(x_j)$ in it's j-th column.

$$\begin{aligned} R(\mathbf{A}, S) &= \frac{1}{m}\sum_{i=1}^{m}\|Y(y_i) - \mathbf{W}X(x_i)\|^2 \\ &= \frac{1}{m}\|\mathbf{M}_{\mathcal{Y}} - \mathbf{W}\mathbf{M}_{\mathcal{X}}\|^2 \\ &= \frac{1}{m}\|\mathbf{M}_{\mathcal{Y}} - \mathbf{M}_{\mathcal{Y}}\mathbf{A}\mathbf{M}_{\mathcal{X}}^{\dagger}\mathbf{M}_{\mathcal{X}}\|^2 \\ &= \frac{1}{m}\|\mathbf{M}_{\mathcal{Y}} - \mathbf{M}_{\mathcal{Y}}\mathbf{A}\mathbf{K}_{\mathcal{X}}\|^2 \\ &= \frac{1}{m}\|\mathbf{M}_{\mathcal{Y}}(\mathbf{I} - \mathbf{A}\mathbf{K}_{\mathcal{X}})\|^2 \end{aligned} \tag{40}$$

$$\begin{aligned} \frac{\partial}{\partial A_{[i;j]}} R(\mathbf{A}, S) &= \frac{1}{m}\frac{\partial}{\partial A_{[i;j]}}\sum_{k,l=1}^{m}\left[\mathbf{M}_{\mathcal{Y}}(\mathbf{I} - \mathbf{A}\mathbf{K}_{\mathcal{X}})\right]_{[k;l]}^{2} \\ &= \frac{2}{m}\sum_{k,l=1}^{m}\left[\mathbf{M}_{\mathcal{Y}}(\mathbf{I} - \mathbf{A}\mathbf{K}_{\mathcal{X}})\right]_{[k;l]}\frac{\partial}{\partial A_{[i;j]}}\left[\mathbf{M}_{\mathcal{Y}}(\mathbf{I} - \mathbf{A}\mathbf{K}_{\mathcal{X}})\right]_{[k;l]} \\ &= \frac{-2}{m}\sum_{k,l=1}^{m}\left[\mathbf{M}_{\mathcal{Y}}(\mathbf{I} - \mathbf{A}\mathbf{K}_{\mathcal{X}})\right]_{[k;l]}\frac{\partial}{\partial A_{[i;j]}}\left[\mathbf{M}_{\mathcal{Y}}\mathbf{A}\mathbf{K}_{\mathcal{X}}\right]_{[k;l]} \\ &= \frac{-2}{m}\sum_{k,l=1}^{m}\left[\mathbf{M}_{\mathcal{Y}}(\mathbf{I} - \mathbf{A}\mathbf{K}_{\mathcal{X}})\right]_{[k;l]}\frac{\partial}{\partial A_{[i;j]}}\left[\sum_{k',l'=1}^{m}\mathbf{M}_{\mathcal{Y}_{[k;k']}}\mathbf{A}_{[k';l']}\mathbf{K}_{\mathcal{X}_{[l';l]}}\right] \\ &= \frac{-2}{m}\sum_{k,l=1}^{m}\left[\mathbf{M}_{\mathcal{Y}}(\mathbf{I} - \mathbf{A}\mathbf{K}_{\mathcal{X}})\right]_{[k;l]}\mathbf{M}_{\mathcal{Y}_{[k;i]}}\mathbf{K}_{\mathcal{X}_{[j;l]}} \\ &= \frac{-2}{m}\sum_{k,l=1}^{m}(\mathbf{M}_{\mathcal{Y}})^{\dagger}_{[i;k]}\left[\mathbf{M}_{\mathcal{Y}}(\mathbf{I} - \mathbf{A}\mathbf{K}_{\mathcal{X}})\right]_{[k;l]}\mathbf{K}_{\mathcal{X}_{[j;l]}} \\ &= \frac{-2}{m}\sum_{l=1}^{m}\left[\mathbf{M}_{\mathcal{Y}}^{\dagger}\mathbf{M}_{\mathcal{Y}}(\mathbf{I} - \mathbf{A}\mathbf{K}_{\mathcal{X}})\right]_{[i;l]}\mathbf{K}_{\mathcal{X}_{[j;l]}} \\ &= \frac{-2}{m}\left[\mathbf{K}_{\mathcal{Y}}(\mathbf{I} - \mathbf{A}\mathbf{K}_{\mathcal{X}})\mathbf{K}_{\mathcal{X}}^{\intercal}\right]_{[i;j]} \\ &= \frac{2}{m}\left[\mathbf{K}_{\mathcal{Y}}(\mathbf{A}\mathbf{K}_{\mathcal{X}} - \mathbf{I})\mathbf{K}_{\mathcal{X}}\right]_{[i;j]} \end{aligned} \tag{41}$$

□

11. Details on how equation (14) becomes $\gamma_{i,j}(\delta_i \lambda_j^2 + m\beta) = \delta_i \lambda_j (u_i^\intercal v_j)$

Because $\{u_i v_j^\intercal\}_{(i,j)\in\mathcal{I}}$ constitutes an orthonormal basis of $\mathbb{R}^{m^2}$ we have

$$\mathbf{A} = \sum_{i=1}^{m}\sum_{j=1}^{m} \gamma_{i,j} u_i v_j^\intercal \tag{42}$$

and the following equalities (recall that $\mathbf{K}_\mathcal{Y} = \sum_{k=1}^m \delta_k u_k u_k^\intercal$ and $\mathbf{K}_\mathcal{X} = \sum_{l=1}^m \lambda_l v_l v_l^\intercal$)

$$\begin{aligned}
\mathbf{K}_\mathcal{Y}\mathbf{K}_\mathcal{X} &= \sum_{k=1}^{m} \delta_k u_k u_k^\intercal \sum_{l=1}^{m} \lambda_l v_l v_l^\intercal \\
&= \sum_{k,l=1}^{m} \delta_k \lambda_l (u_k^\intercal v_l) u_k v_l^\intercal \\
\mathbf{K}_\mathcal{X}^2 &= \sum_{l=1}^{m} \lambda_l v_l v_l^\intercal \sum_{l'=1}^{m} \lambda_{l'} v_{l'} v_{l'}^\intercal \\
&= \sum_{l=1}^{m} \lambda_l^2 v_l v_l^\intercal \\
\mathbf{A}\mathbf{K}_\mathcal{X}^2 &= \sum_{k=1}^{m}\sum_{l=1}^{m} \gamma_{k,l} u_k v_l^\intercal \sum_{l=1}^{m} \lambda_l^2 v_l v_l^\intercal \\
&= \sum_{k=1}^{m}\sum_{l=1}^{m} \gamma_{k,l} \lambda_l^2 u_k v_l^\intercal \\
\mathbf{K}_\mathcal{Y}\mathbf{A}\mathbf{K}_\mathcal{X}^2 &= \sum_{k'=1}^{m} \delta_{k'} u_{k'} u_{k'}^\intercal \sum_{k=1}^{m}\sum_{l=1}^{m} \gamma_{k,l} \lambda_l^2 u_k v_l^\intercal \\
&= \sum_{k=1}^{m}\sum_{l=1}^{m} \gamma_{k,l} \delta_k \lambda_l^2 u_k v_l^\intercal
\end{aligned}$$

Equation (14) then becomes

$$\begin{aligned}
\frac{2}{m}\mathbf{K}_\mathcal{Y}(\mathbf{A}\mathbf{K}_\mathcal{X} - \mathbf{I})\mathbf{K}_\mathcal{X} + 2\beta\mathbf{A} &= 0 \\
\frac{2}{m}\mathbf{K}_\mathcal{Y}\mathbf{A}\mathbf{K}_\mathcal{X}^2 - \frac{2}{m}\mathbf{K}_\mathcal{Y}\mathbf{K}_\mathcal{X} + 2\beta\mathbf{A} &= 0 \\
\sum_{k=1}^{m}\sum_{l=1}^{m}\left[\frac{2}{m}\gamma_{k,l}\delta_k\lambda_l^2 - \frac{2}{m}\lambda_l(u_k^\intercal v_l) + 2\beta\gamma_{k,l}\right] u_k v_l^\intercal &= 0
\end{aligned}$$

Since $u_k v_l^\intercal$ are linearly independent vectors of $\mathbb{R}^{m^2}$, the previous equation is satisfied when

$$\begin{aligned}
\frac{2}{m}\gamma_{k,l}\delta_k\lambda_l^2 - \frac{2}{m}\lambda_l(u_k^\intercal v_l) + 2\beta\gamma_{k,l} &= 0 \\
\frac{2}{m}\gamma_{k,l}\delta_k\lambda_l^2 + 2\beta\gamma_{k,l} &= \frac{2}{m}\lambda_l(u_k^\intercal v_l) \\
\gamma_{k,l}(\delta_k\lambda_l^2 + m\beta) &= \delta_k\lambda_l(u_k^\intercal v_l)
\end{aligned}$$