Supplemental Material: ANALYZING CROSS VALIDATION IN COMPRESSED SENSING WITH MIXED GAUSSIAN AND IMPULSE MEASUREMENT NOISE WITH L1 ERRORS

Shubhang Bhatnagar[♯*], Chinmay Gurjarpadhye [♯*] and Ajit Rajwade[†*]

1 Introduction

In this document, we present the proofs of Lemma 1, Theorem 1, Lemma 2 and Theorem 2 from the main paper.

Assumption 1: As discussed in the problem formulation, we assume the mean of the Gaussian noise to be large when compared to the values in signal $\boldsymbol{x}$ and to the Gaussian noise variance which translates to the following approximation,

$$|\sum_{j=1}^{N} A_{\mathrm{cv},i,j}\Delta x_j + n_i + G_i| \sim \sum_{j=1}^{N} A_{\mathrm{cv},i,j}\Delta x_j + n_i + G_i \text{ and}$$
$$|\sum_{j=1}^{N} A_{\mathrm{cv},i,j}\Delta x_j + n_i - G_i| \sim \sum_{j=1}^{N} -A_{\mathrm{cv},i,j}\Delta x_j - n_i + G_i$$

Remark 1: Throughout these proofs, we come across the folded Gaussian distribution, i.e. the distribution of $|X|$ if $X \sim \mathcal{N}(\mu, \sigma^2)$. We note that that the mean, μ' of this distribution is given as $\mu' = \sigma\sqrt{\frac{2}{\pi}}e^{-\mu^2/2\sigma^2} - \mu\left(1 - 2\Phi\left(\frac{\mu}{\sigma}\right)\right)$, where $\Phi(\cdot)$ is the cdf of the Gaussian distribution. This is proved in [1, 2].
Also note that all the expectations in the proofs that follow are over the noise instances as well as the instances for the randomly generated sensing matrix $\boldsymbol{A}$.

2 Proof of Lemma 1

The L1 cross validation error ϵ_{cv} (shorthand for $\epsilon_{cv,\ell_1,\lambda}$ in the main paper) is given as follows,

$$\epsilon_{\mathrm{cv}} = \|\boldsymbol{y}_{\mathrm{cv}} - \boldsymbol{\Phi_{cv}}\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}\|_1 = \sum_{i=1}^{m_{\mathrm{cv}}} \left|(\sum_{j=1}^{N} A_{\mathrm{cv},i,j}\Delta x_j) + n_i + B_i G_i\right|,$$

*♯- CG and SB are first authors with equal contribution. AR acknowledges support from SERB Grant #10013890.

where $\boldsymbol{x}$ is the true, unknown signal, $\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}$ is its estimate and $\Delta x_j := x_j - \hat{x}_\lambda(j)$. For simplicity of notation we define r_i as follows,

$$r_i = \left| \sum_{j=1}^{N} A_{\text{cv},i,j} \Delta x_j + n_i + B_i G_i \right|$$

Using this definition we have,

$$\epsilon_{\text{cv}} = \sum_{i=1}^{m_{\text{cv}}} r_i$$

Now we compute the value of $E[r_i]$,

$$E[r_i] = \frac{b}{2} E[r_i \mid B_i = -1] + (1-b) E[r_i \mid B_i = 0] + \frac{b}{2} E[r_i \mid B_i = 1]. \tag{1}$$

We first compute the value of middle term in (1):

$$E[r_i \mid B_i = 0] = E\left[|\sum_{j=1}^{N} A_{\text{cv},i,j} \Delta x_j + n_i| \right].$$

Here observe that the random variable $(\sum_{j=1}^{N} A_{\text{cv},i,j} \Delta x_j + n_i)$ has mean 0 and its variance can be computed as follows,

$$E\left[(\sum_{j=1}^{N} A_{\text{cv},i,j} \Delta x_j + n_i)^2 \right] = \sum_{j=1}^{N} \Delta x_j^2 E(A_{\text{cv},i,j}^2) + E(n_i^2) = \frac{1}{m} \left(\sum_{i}^{N} \Delta x_j^2 + \sigma_n^2 \right).$$

Using the formula for the mean of a folded Gaussian, we have:

$$E(r_i \mid B_i = 0) = \sqrt{\frac{2(\sum_{j=1}^{N} \Delta x_j^2 + \sigma_n^2)}{\pi m}}.$$

Next, we calculate the conditional expectation for the other 2 terms which have $B = \pm 1$-

$$E(r_i \mid B_i = \pm 1) = E(|\sum_{j=1}^{N} a_{\text{cv},i,j} \Delta x_j + n_i \pm G_i|).$$

We now use Assumption 1 to approximate $|\sum_{j=1}^{N} A_{\text{cv},i,j} \Delta x_j + n_i + G_i| \sim \sum_{j=1}^{N} A_{\text{cv},i,j} \Delta x_j + n_i + G_i$ and $|\sum_{j=1}^{N} A_{\text{cv},i,j} \Delta x_j + n_i - G_i| \sim \sum_{j=1}^{N} -A_{\text{cv},i,j} \Delta x_j - n_i + G_i$. Now $\sum_{j=1}^{N} A_{\text{cv},i,j} \Delta x_j + n_i + G_i$ is a sum of independent Gaussian Random variables, and hence is a Gaussian with the following mean:

$$E(A_{\text{cv},i,j} \Delta x_j) + E(n_i) + E(G_i) = \mu_G.$$

Similarly,$\sum_{j=1}^{N} -A_{\text{cv},i,j}\Delta x_j - n_i + G_i$ is also a sum of independent Gaussian Random variables, and hence is a Gaussian with mean

$$E(A_{\text{cv},i,j}\Delta x_j) - E(n_i) + E(G_i) = \mu_G.$$

So, we have

$$E(r_i \mid B_i = \pm 1) = \mu_G.$$

We take $\varepsilon_x^2 := \sum_{j=1}^{N} \Delta x_j^2$ which is the actual recovery error squared given the signal $\boldsymbol{x}$.

Using these and substituting in (1) we get the mean of r_i as-

$$\begin{aligned} E(r_i) =& \frac{b}{2}E(r_i \mid B_i = -1) + (1-b)E(r_i \mid B_i = -1) + \frac{b}{2}E(|r_i| \mid B_i = 1) \\ =& \; b\mu_G + (1-b)\sqrt{\frac{2}{m\pi}\left(\varepsilon_x^2 + \sigma_n^2\right)}. \end{aligned}$$

Next, we calculate the variance of r_i

$$Var(r_i) = E(r_i^2) - E(r_i)^2. \tag{2}$$

We first calculate $E(r_i^2)$:

$$E(r_i^2) = \sum_{j=1}^{N} E(A_{\text{cv},i,j}^2)\Delta x_j^2 + E(N_{\text{cv},i}^2) + E(B_i^2 G_i^2) + 2E(B_i G_i N_{\text{cv},i}) + \tag{3}$$

$$2\sum_{j=1}^{N}\left(E(A_{\text{cv},i,j}N_{\text{cv},i})\Delta x_j\right) + 2\sum_{j=1}^{N}\left(E(A_{\text{cv},i,j}B_i G_i)\Delta x_j\right)$$

$$\stackrel{(a)}{=} \frac{\varepsilon_x^2}{m} + \frac{\sigma_n^2}{m} + E(B_i^2)E(G_i^2) + 0 + 0 + 0 \tag{4}$$

$$= \frac{\varepsilon_x^2 + \sigma_n^2}{m} + b(\sigma_G^2 + \mu_G^2). \tag{5}$$

Here equality marked as '(a)' follows because $N_{\text{cv},i}, B_i, G_i$ and $A_{\text{cv},i,j}$ are all independent of each other and $N_{\text{cv},i}, B_i$ and $A_{\text{cv},i,j}$ have 0 mean.
Now calculating the term $E(r_i)^2$:

$$E(r_i)^2 = b^2\mu_G^2 + (1-b)^2\frac{2}{m\pi}(\sigma_n^2 + \varepsilon_x^2) + 2b(1-b)\mu_G\sqrt{\frac{2}{m\pi}(\sigma_n^2 + \varepsilon_x^2)}$$

Substituting these terms in equation (2), we obtain $\text{Var}(r_i)$ as follows:

$$\text{Var}(r_i) = \left(1 - (1-b)^2\frac{2}{\pi}\right)\frac{\varepsilon_x^2 + \sigma_n^2}{m} + (b)(\sigma_G^2 + (1-b)\mu_G^2) - 2b(1-b)\mu_G\sqrt{\frac{2}{m\pi}(\sigma_n^2 + \varepsilon_x^2)}.$$

Now seeing that $\epsilon_{\text{cv}} = \sum_{i=1}^{m_{\text{cv}}} |r_i|$, and assuming that m_{cv} is large, we apply the Central limit theorem (CLT) to say that ϵ_{cv} follows a normal distribution with the following mean and variance-

$$\epsilon_{cv} \sim N(m_{\text{cv}}E(r_i), m_{\text{cv}}Var(r_i))$$

with -

$$\begin{aligned}
E(r_i) &= b\mu_G + (1-b)\sqrt{\frac{2}{m\pi}\left(\varepsilon_x^2+\sigma_n^2\right)} \\
\mathrm{Var}(r_i) &= \left(1-(1-b)^2\frac{2}{\pi}\right)\frac{\varepsilon_x^2+\sigma_n^2}{m} + (b)(\sigma_G^2+(1-b)\mu_G^2) - 2b(1-b)\mu_G\sqrt{\frac{2}{m\pi}(\sigma_n^2+\varepsilon_x^2)}.
\end{aligned}$$

This completes the proof of Lemma 1. ■

3 Proof of Theorem 1

Now assuming $\mu_1 = E(\epsilon_{\text{cv}})$ and the standard deviation of ϵ_{cv} as σ_1, we can say that with probability $\mathrm{erf}(\frac{\varrho}{\sqrt{2}})$ the following inequality holds:

$$-\varrho \leq \frac{\epsilon_{\text{cv}} - \mu_1}{\sigma_1} \leq \varrho. \tag{6}$$

Consider $K_1 := \left(1-(1-b)^2\frac{2}{\pi}\right), K_2 := b\left(\sigma_G^2+(1-b)\mu_G^2\right)$. Consider the expression for the variance of ϵ_{cv} in Lemma 1. Using the inequality $\sqrt{a}+\sqrt{b} \geq \sqrt{a+b}$ where $a, b > 0$, the following inequalities hold:

$$\sqrt{\frac{m_{\text{cv}}}{m}K_1(\varepsilon_x^2+\sigma_n^2)} + \sqrt{m_{\text{cv}}K_2} \geq \sqrt{\frac{m_{\text{cv}}}{m}K_1(\varepsilon_x^2+\sigma_n^2) + m_{\text{cv}}K_2} \tag{7}$$

This is because $K_1, K_2, m, m_{\text{cv}}$ are all positive quantities.
Consider. $\sigma_2 = \sqrt{\frac{m_{\text{cv}}}{m}K_1\varepsilon_x^2+\sigma_n^2} + \sqrt{m_{\text{cv}}K_2}$. From inequality (7) we can say that $\sigma_2 \geq \sigma_1$

Now using inequality (6) and (7) we arrive at the following inequality:

$$\begin{aligned}
-\varrho &\leq \frac{\epsilon_{\text{cv}}-\mu_1}{\sigma_2} \leq \varrho \\
-\sigma_2\varrho &\leq \epsilon_{\text{cv}} - \mu_1 \leq \sigma_2\varrho \\
\mu_1 - \sigma_2\varrho &\leq \epsilon_{\text{cv}} \leq \mu_1 + \sigma_2\varrho
\end{aligned} \tag{8}$$

Now looking at the LHS of the above inequality, we get-

$$\begin{aligned}
-\varrho\sqrt{\frac{m_{\text{cv}}K_1}{m}(\varepsilon_x^2+\sigma_n^2)} - \varrho\sqrt{m_{\text{cv}}K_2} + m_{\text{cv}}b\mu_G + (1-b)m_{\text{cv}}\sqrt{\frac{2}{m\pi}\varepsilon_x^2+\sigma_n^2} &\leq \epsilon_{\text{cv}} \\
\sqrt{\varepsilon_x^2+\sigma_n^2}\left((1-b)m_{\text{cv}}\sqrt{\frac{2}{m\pi}} - \varrho\sqrt{\frac{m_{\text{cv}}K_1}{m}}\right) &\leq \epsilon_{\text{cv}} + \varrho\sqrt{m_{\text{cv}}K_2} - m_{\text{cv}}b\mu_G \\
\sqrt{\varepsilon_x^2+\sigma_n^2} \leq \frac{\sqrt{m}}{m_{\text{cv}}}\left(\frac{\epsilon_{\text{cv}}+\varrho\sqrt{m_{\text{cv}}K_2}-m_{\text{cv}}b\mu_G}{\left((1-b)\sqrt{\frac{2}{\pi}} - \varrho\sqrt{\frac{K_1}{m_{\text{cv}}}}\right)}\right)
\end{aligned} \tag{9}$$

Now seeing RHS of the inequality (8)

$$
\begin{aligned}
\epsilon_{\mathrm{cv}} &\leq \varrho\sqrt{\frac{m_{\mathrm{cv}}K_1}{m}(\varepsilon_x^2+\sigma_n^2)}+\varrho\sqrt{m_{\mathrm{cv}}K_2}+m_{\mathrm{cv}}b\mu_G+(1-b)m_{\mathrm{cv}}\sqrt{\frac{2}{m\pi}(\varepsilon_x^2+\sigma_n^2)} \\
\epsilon_{\mathrm{cv}}-\varrho\sqrt{m_{\mathrm{cv}}K_2}-m_{\mathrm{cv}}b\mu_G &\leq \sqrt{\varepsilon_x^2+\sigma_n^2}\left(\varrho\sqrt{\frac{m_{\mathrm{cv}}K_1}{m}}+(1-b)m_{\mathrm{cv}}\sqrt{\frac{2}{m\pi}}\right) \\
\sqrt{\varepsilon_x^2+\sigma_n^2} &\geq \frac{\sqrt{m}}{m_{\mathrm{cv}}}\left(\frac{\epsilon_{\mathrm{cv}}-\varrho\sqrt{m_{\mathrm{cv}}K_2}-m_{\mathrm{cv}}b\mu_G}{\varrho\sqrt{\frac{K_1}{m_{\mathrm{cv}}}}+(1-b)\sqrt{\frac{2}{\pi}}}\right).
\end{aligned}
$$

Defining

$$
\begin{aligned}
p(\varrho,\pm) &:= m_{\mathrm{cv}}b\mu_G \pm \varrho\sqrt{m_{\mathrm{cv}}K_2} \\
h(\varrho,\pm) &:= (1-b)\sqrt{\frac{2}{\pi}} \pm \varrho\sqrt{\frac{K_1}{m_{\mathrm{cv}}}},
\end{aligned}
$$

We arrive at the following confidence interval on ε_x (with probability $\mathrm{erf}(\varrho/\sqrt{2})$):

$$
\frac{\sqrt{m}}{m_{\mathrm{cv}}}\frac{\epsilon_{\mathrm{cv}}-p(\varrho,+)}{h(\varrho,+)} \leq \sqrt{\varepsilon_x+\sigma_n^2} \leq \frac{\sqrt{m}}{m_{\mathrm{cv}}}\frac{\epsilon_{\mathrm{cv}}-p(\varrho,-)}{h(\varrho,-)}. \tag{10}
$$

This proves Theorem 1. ■

We observe that the confidence interval length tends to 0 as we increase the value of m_{cv}, as expected. We prove this claim in the following steps-

$$
\begin{aligned}
&\frac{\sqrt{m}}{m_{\mathrm{cv}}}\left(\frac{\epsilon_{\mathrm{cv}}-p(\varrho,-)}{h(\varrho,-)}-\right)\frac{\sqrt{m}}{m_{\mathrm{cv}}}\left(\frac{\epsilon_{\mathrm{cv}}-p(\varrho,+)}{h(\varrho,+)}\right) \\
&=\frac{\sqrt{m}}{m_{\mathrm{cv}}}\left(\frac{\epsilon_{\mathrm{cv}}(h(\varrho,+)-h(\varrho,+))-h(\varrho,+)p(\varrho,-)+h(\varrho,-)p(\varrho,+)}{h(\varrho,-)h(\varrho,+)}\right) \\
&=\frac{\sqrt{m}}{m_{\mathrm{cv}}}\left(\frac{2\varrho\epsilon_{\mathrm{cv}}\sqrt{\frac{K_1}{m_{\mathrm{cv}}}}+2(1-b)\sqrt{\frac{2}{\pi}}\varrho\sqrt{K_2 m_{\mathrm{cv}}}-2\sqrt{\frac{K_1}{m_{\mathrm{cv}}}}b\mu_G\varrho m_{\mathrm{cv}}}{(1-b)^2\frac{2}{\pi}-\varrho^2\frac{K_1}{m_{\mathrm{cv}}}}\right) \\
&=\sqrt{m}\left(\frac{2\varrho\epsilon_{\mathrm{cv}}\sqrt{\frac{K_1}{m_{\mathrm{cv}}^3}}+2(1-b)\sqrt{\frac{2}{\pi}}\frac{\varrho}{\sqrt{m_{\mathrm{cv}}}}\sqrt{K_2}-2\sqrt{\frac{K_1}{m_{\mathrm{cv}}}}b\mu_G\varrho}{(1-b)^2\frac{2}{\pi}-\varrho^2\frac{K_1}{m_{\mathrm{cv}}}}\right)
\end{aligned}
$$

Taking limit $m_{\mathrm{cv}} \to \infty$ we get-

$$
\begin{aligned}
&=m\left(\frac{\lim_{m_{\mathrm{cv}}\to\infty}2\varrho\epsilon_{\mathrm{cv}}\sqrt{\frac{K_1}{m_{\mathrm{cv}}^3}}+\lim_{m_{\mathrm{cv}}\to\infty}2(1-b)\sqrt{\frac{2K_2}{\pi m_{\mathrm{cv}}}}\varrho-\lim_{m_{\mathrm{cv}}\to\infty}2\sqrt{\frac{K_1}{m_{\mathrm{cv}}}}b\mu_G\varrho}{\lim_{m_{\mathrm{cv}}\to\infty}2(1-b)\sqrt{\frac{2K_2}{\pi m_{\mathrm{cv}}}}\frac{\varrho}{\sqrt{m_{\mathrm{cv}}}}\sqrt{K_2}-\lim_{m_{\mathrm{cv}}\to\infty}\varrho^2\frac{K_1}{m_{\mathrm{cv}}}}\right) \\
&=m\left(\frac{0+0+0}{(1-b)^2\frac{2}{\pi}+0}\right) \\
&=0
\end{aligned}
$$

Hence, we say that the length of our confidence interval tends to 0 as we increase the value of m_{cv}.

4 Proof of Lemma 2

Here, we wish to obtain the distribution of $\Delta\epsilon_{cv}$. We have:

$$\begin{aligned}\Delta\epsilon_{\text{cv}} &= \epsilon^p_{\text{cv}} - \epsilon^q_{\text{cv}} \\ &= \Sigma_{i=1}^{m_{\text{cv}}} \left(|\Sigma_{j=1}^N A_{\text{cv},ij}\Delta x_j^p + n_{cv,i} + B_{\text{cv},i}G_{\text{cv},i}| - |\Sigma_{j=1}^N A_{\text{cv},ij}\Delta x_j^q + n_{cv,i} + B_{\text{cv},i}G_{\text{cv},i}| \right).\end{aligned}$$

For simplicity we define r_i^p, r_i^q and r_i as follows,

$$\begin{aligned}&r_i^p := |\Sigma_{j=1}^N A_{\text{cv},ij}\Delta x_j^p + n_{cv,i} + B_{\text{cv},i}G_{\text{cv},i}| \qquad r_i^q := |\Sigma_{j=1}^N A_{\text{cv},ij}\Delta x_j^q + n_{cv,i} + B_{\text{cv},i}G_{\text{cv},i}| \\ &r_i := r_i^p - r_i^q.\end{aligned}$$

We now compute $E[r_i]$. To that end, we compute $E[r_i^p]$.

$$E[r_i^p] = (1-b)E[r_i^p|B_i = 0] + \frac{b}{2}E[r_i^p|B_i = 1] + \frac{b}{2}E[r_i^p|B_i = -1]. \tag{11}$$

We expand each of the above terms. We first compute $E[r_i^p|B_i = 1]$ as follows,

$$\begin{aligned}E[r_i^p|B_i = 1] &= E[|\Sigma_{j=1}^N A_{\text{cv},ij}\Delta x_j^p + n_{cv,i} + B_{\text{cv},i}G_{\text{cv},i}||B_i = 1] \\ &= E[|\Sigma_{j=1}^N A_{\text{cv},ij}\Delta x_j^p + n_{cv,i} + G_{\text{cv},i}|] \\ &\overset{(a)}{=} E[\Sigma_{j=1}^N A_{\text{cv},ij}\Delta x_j^p + n_{cv,i} + G_{\text{cv},i}] \\ &= \mu_g.\end{aligned}$$

Here (a) follows from Assumption 1. Similarly, one can get $E[r_i^p|B_i = -1] = \mu_g$. Now, we compute the first term in (11).

$$\begin{aligned}E[r_i^p|B_i = 0] &= E[|\Sigma_{j=1}^N A_{\text{cv},ij}\Delta x_j^p + n_{cv,i} + B_{\text{cv},i}G_{\text{cv},i}|B_i = 0] \\ &= E[|\Sigma_{j=1}^N A_{\text{cv},ij}\Delta x_j^p + n_{cv,i}|] \\ &= E[|\Sigma_{j=1}^N A_{\text{cv},ij}\Delta x_j^p + n_{cv,i}|].\end{aligned}$$

Observe that $\Sigma_{j=1}^N A_{\text{cv},ij}\Delta x_j^p + n_{cv,i}$ is a Gaussian with mean 0 and variance $\frac{\varepsilon_p^2+\sigma_n^2}{m}$, where $\varepsilon_p^2 = \sum_{j=1}^N \Delta x_j^2$. Thus the absolute value is a folded Gaussian and using Remark 1 we get,

$$E[r_i^p|B_i = 0] = \sqrt{\frac{2}{\pi}}\sqrt{\frac{\varepsilon_p^2 + \sigma_n^2}{m}}.$$

For simplicity, here onward we use K_1 and σ_p to denote $\sqrt{\frac{2}{\pi}}$ and $\sqrt{\frac{\varepsilon_p^2+\sigma_n^2}{m}}$ respectively. Similarly, we denote $\sqrt{\frac{\varepsilon_q^2+\sigma_n^2}{m}}$ by σ_q. Thus, continuing from (11) we get,

$$\begin{aligned}E[r_i^p] &= (1-b)E[r_i^p|B_i = 0] + \frac{b}{2}E[r_i^p|B_i = 1] + \frac{b}{2}E[r_i^p|B_i = -1] \\ &= (1-b)\sqrt{\frac{2}{\pi}}\sqrt{\frac{\varepsilon_p^2 + \sigma_n^2}{m}} + \mu_g - \mu_g \\ &= (1-b)K_1\sigma_p.\end{aligned}$$

Due to the symmetry we will have,

$$E[r_i^q] = (1-b)K_1\sigma_q$$

Thus from the definition of r_i we get,

$$E[r_i] = (1-b)K_1(\sigma_p - \sigma_q) \tag{12}$$

Having computed $E[r_i]$ we now compute $E[r_i^2]$ as follows,

$$E[r_i^2] = (1-b)E[r_i^2|B_i=0] + \frac{b}{2}E[r_i^2|B_i=1] + \frac{b}{2}E[r_i^2|B_i=-1] \tag{13}$$

We expand each of the above terms. We first compute $E[r_i^2|B_i=1]$ as follows,

$$\begin{aligned}
&E[r_i^2|B_i=1] \\
&= E\left[\left(|\Sigma_{j=1}^N A_{\mathrm{cv},ij}\Delta x_j^p + n_{cv,i} + B_{\mathrm{cv},i}G_{\mathrm{cv},i}| - |\Sigma_{j=1}^N A_{\mathrm{cv},ij}\Delta x_j^q + n_{cv,i} + B_{\mathrm{cv},i}G_{\mathrm{cv},i}|\right)^2 |B_i=1\right] \\
&= E\left[\left(|\Sigma_{j=1}^N A_{\mathrm{cv},ij}\Delta x_j^p + n_{cv,i} + G_{\mathrm{cv},i}| - |\Sigma_{j=1}^N A_{\mathrm{cv},ij}\Delta x_j^q + n_{cv,i} + G_{\mathrm{cv},i}|\right)^2\right] \\
&\overset{(a)}{=} E\left[\left(\Sigma_{j=1}^N A_{\mathrm{cv},ij}(\Delta x_j^p - \Delta x_j^q)\right)^2\right] \\
&= \frac{1}{m}\left(\varepsilon_p^2 + \varepsilon_p^2 - 2\langle \Delta\mathbf{x}^p, \Delta\mathbf{x}^q\rangle\right)
\end{aligned} \tag{14}$$

Here, (a) follows from assumption 1. Similarly we get,

$$E[r_i^2|B_i=-1] = \frac{1}{m}\left(\varepsilon_p^2 + \varepsilon_q^2 - 2\langle \Delta\mathbf{x}^p, \Delta\mathbf{x}^q\rangle\right) \tag{15}$$

Now we compute the first term in (13) as follows,

$$\begin{aligned}
&E[r_i^2|B_i=0] \\
&= E\Bigg[|\Sigma_{j=1}^N A_{\mathrm{cv},ij}\Delta x_j^p + n_{cv,i}|^2 + |\Sigma_{j=1}^N A_{\mathrm{cv},ij}\Delta x_j^q + n_{cv,i}|^2 \\
&\quad - 2|\Sigma_{j=1}^N A_{\mathrm{cv},ij}\Delta x_j^p + n_{cv,i}||\Sigma_{j=1}^N A_{\mathrm{cv},ij}\Delta x_j^q + n_{cv,i}|\Bigg]
\end{aligned} \tag{16}$$

Recall that $\Sigma_{j=1}^N A_{\mathrm{cv},ij}\Delta x_j^p + n_{cv,i} \sim \mathcal{N}(0,\sigma_p^2)$ as argued earlier. Similarly we have $\Sigma_{j=1}^N A_{\mathrm{cv},ij}\Delta x_j^q + n_{cv,i} \sim \mathcal{N}(0,\sigma_q^2)$. So we now compute the remaining term in (16). For simplicity we define Gaussian random variables X and Y as follows,

$$X = \Sigma_{j=1}^N A_{\mathrm{cv},ij}\Delta x_j^p + n_{cv,i} \qquad Y = \Sigma_{j=1}^N A_{\mathrm{cv},ij}\Delta x_j^q + n_{cv,i}$$

Note that X and Y are 0 mean but correlated random variables. We further define random variables X' and Y' as follows,

$$X' = X/\sigma_p \qquad Y' = Y/\sigma_q$$

Notice that X' and Y' are normally distributed. Let ρ be the covariance of X' and Y'. We compute the value of ρ later. We can write X' and Y' as follows,

$$X' = U \qquad Y' = \rho U + \sqrt{1-\rho^2}V$$

where U and V are independent and identically distributed Gaussian random variables, with mean 0 and variance 1. Observe that $E[|XY|] = \sigma_p\sigma_q E[|X'Y'|]$. We now compute $E[|X'Y'|]$ as follows,

$$\begin{aligned}
&E[|X'Y'|]\\
&= E[|U(\rho U + \sqrt{1-\rho^2}V)|]\\
&= \sqrt{1-\rho^2}E[|U(K_3U+V)|] \qquad \left(\text{Here } K_3 := \frac{\rho}{\sqrt{1-\rho^2}}\right)\\
&= \sqrt{1-\rho^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|K_3x^2+xy|f_U(x)f_V(y)dydx\\
&= \sqrt{1-\rho^2}\int_{-\infty}^{\infty}f_U(x)|x|\left(\int_{-\infty}^{\infty}|K_3x+y|f_V(y)dy\right)dx\\
&\overset{(a)}{=} \sqrt{1-\rho^2}\int_{-\infty}^{\infty}f_U(x)|x|\left(K_1e^{\frac{-x^2K_3^2}{2}}+K_3x\left(\text{erf}\left(\frac{K_3x}{\sqrt{2}}\right)\right)\right)dx\\
&= \sqrt{1-\rho^2}\int_{-\infty}^{\infty}f_U(x)|x|K_1e^{\frac{-x^2K_2^2}{2}}+\int_{-\infty}^{\infty}f_U(x)|x|K_3x\left(\text{erf}\left(\frac{K_3x}{\sqrt{2}}\right)\right)dx\\
&= \sqrt{1-\rho^2}\int_{-\infty}^{\infty}\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}|x|K_1e^{\frac{-x^2K_2^2}{2}}+\int_{0}^{\infty}2f_U(x)|x|K_3x\left(\text{erf}\left(\frac{K_3x}{\sqrt{2}}\right)\right)dx\\
&\overset{(b)}{=} \sqrt{1-\rho^2}\left(\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}|x|K_1e^{\frac{-x^2(K_3^2+1)}{2}}+\frac{2K_3}{\sqrt{2\pi}}\left[\frac{\sqrt{\pi}}{4}\times 2\sqrt{2}-\frac{1}{2\sqrt{\pi}}\left(2\sqrt{2}\tan^{-1}\left(\frac{1}{K_3}\right)-\frac{(K_3/\sqrt{2})}{\frac{1}{2}(1/2+K_3^2/2)}\right)\right]\right)\\
&= \sqrt{1-\rho^2}\left(\sqrt{\frac{2}{\pi}}\sqrt{\frac{2}{\pi}}\times\frac{1}{K_3^2+1}+K_3\left[1-\frac{2}{\pi}\tan^{-1}\left(\frac{1}{K_3}\right)+\frac{2}{\pi}\frac{K_3}{K_3^2+1}\right]\right)\\
&= \sqrt{1-\rho^2}\left(\frac{2}{\pi}\frac{1}{K_3^2+1}+\frac{\rho}{\sqrt{1-\rho^2}}\left[1-\frac{2}{\pi}\tan^{-1}\left(\frac{\sqrt{1-\rho^2}}{\rho}\right)+\frac{2\rho}{\pi}\sqrt{1-\rho^2}\right]\right)\\
&= \rho-\frac{2\rho\tan^{-1}(\sqrt{1-\rho^2}/\rho)}{\pi}+\frac{2\sqrt{1-\rho^2}}{\pi}.
\end{aligned}$$

Here, (a) follows from Remark 1; (b) follows from **??**. Now we compute the value of ρ as follows:

$$\begin{aligned}
\rho = E[X'Y'] &= \frac{E[XY]}{\sigma^p\sigma^q}\\
&= \frac{E[N^2_{\text{cv},ij}+\Sigma_{j=1}^N A^2_{\text{cv},ij}\Delta x_j^p\Delta x_j^q]}{\sigma^p\sigma^q}\\
&= \frac{1}{\sigma^p\sigma^q}\left(\frac{\sigma_N^2}{m}+\frac{\langle\Delta x_j^p\Delta x_j^q\rangle}{m}\right)
\end{aligned}\tag{17}$$

Now, substituting all terms in (13) we get,

$$E[r_i^2] = (1-b)\left(\sigma_p^2+\sigma_q^2-2\rho_1\sigma_p\sigma_q\right)+\frac{b}{m}\left(\varepsilon_p^2+\varepsilon_p^2-2\langle\Delta x_j^p,\Delta x_j^q\rangle\right)$$

where

$$\rho_1 = \rho - \frac{2\rho\tan^{-1}(\sqrt{1-\rho^2}/\rho)}{\pi} + \frac{2\sqrt{1-\rho^2}}{\pi}$$

Here, ρ is as given in (17). Also $mean(r_i) = E[r_i]$ is known from (12). Thus we have the variance of r_i as,

$$\begin{aligned} Var(r_i) &= E[r_i^2] - (E[r_i])^2 \\ &= (1-b)\left(\sigma_p^2 + \sigma_q^2 - 2\rho_1\sigma_p\sigma_q\right) + \frac{b}{m}\left(\varepsilon_p^2 + \varepsilon_p^2 - 2\langle \Delta x_j^p, \Delta x_j^q\rangle\right) - ((1-b)K_1(\sigma_p - \sigma_q))^2 \end{aligned}$$

Having obtained the mean and variance of r_i, we now use the Central Limit Theorem to compute the distribution of $\Delta\epsilon_{\text{cv}}$.

$$\Delta\epsilon_{\text{cv}} \sim \mathcal{N}(\mu, \sigma^2) \sim \mathcal{N}(m_{\text{cv}} mean(r_i), m_{\text{cv}} Var(r_i))$$

This completes the proof of Lemma 2. ■

5 Proof of Theorem 2

Theorem 2 follows directly from Lemma 2.

$$\begin{aligned} Pr(\epsilon_{cv}^p > \epsilon_{cv}^q) &= Pr(\Delta\epsilon_{\text{cv}} > 0) \\ &= \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx \end{aligned}$$

Here, μ and σ are as given in Lemma 2. Now substituting $x = (t\sigma + \mu)$, gives

$$\begin{aligned} Pr(\epsilon_{cv}^p > \epsilon_{cv}^q) &= \int_{-\mu/\sigma}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dx \\ &= \int_{-\infty}^{\mu/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dx \\ &= \Phi(\frac{\mu}{\sigma}) \end{aligned} \tag{18}$$

This completes the proof of Theorem 2. ■

References

[1] "Folded normal distribution," https://en.wikipedia.org/wiki/Folded$_normal_distribution.$

[2] F. C. Leone, L. S. Nelson, and R. B. Nottingham, "The folded normal distribution," *Technometrics*, vol. 3, no. 4, pp. 543–550, 1961.