Appendix A. Proof of Theorem 1

Expanding the objective of (2) for DS, we obtain

$$\begin{aligned}\mathbb{E}_{P_{X,Y,D}}[\ell(f(X),Y)] &= \mathbb{E}_{P_{Y,D}}\mathbb{E}_{P_{X|Y,D}}[\ell(f(X),Y)|Y,D] \\ &= \sum_{(y,d)\in\mathcal{Y}\times\mathcal{D}} \pi^{(y,d)}\mathbb{E}_{P_{X|Y,D}}[\ell(f(X),Y)|Y=y,D=d] \\ &= \frac{1}{4}\sum_{(y,d)\in\mathcal{Y}\times\mathcal{D}} \mathbb{E}_{P_{X|Y,D}}[\ell(f(X),Y)|Y=y,D=d]. \end{aligned} \tag{8}$$

Expanding the objective of (2) for UW, we obtain

$$\begin{aligned}\mathbb{E}_{P_{X,Y,D}}[\ell(f(X),Y)c(Y,D)] &= \mathbb{E}_{P_{Y,D}}\left[c(Y,D)\mathbb{E}_{P_{X|Y,D}}[\ell(f(X),Y)|Y,D]\right] \\ &= \sum_{(y,d)\in\mathcal{Y}\times\mathcal{D}} \frac{\pi^{(y,d)}}{4\pi^{(y,d)}}\mathbb{E}_{P_{X|Y,D}}[\ell(f(X),Y)|Y=y,D=d] \\ &= \frac{1}{4}\sum_{(y,d)\in\mathcal{Y}\times\mathcal{D}} \mathbb{E}_{P_{X|Y,D}}[\ell(f(X),Y)|Y=y,D=d]. \end{aligned} \tag{9}$$

Therefore, the objectives of DS and UW are the same, and as a result so are the corresponding minimizers, if they exist.

Appendix B. Proof of Theorem 3

Our proof can be outlined as involving three steps; these steps rely on Theorem 1 and include one new lemma. We enumerate the steps below:

1. We first show in Theorem 5 that for any minority group prior π_0, the WGA for SRM is given by (10).
2. Using the WGA for SRM derived in Theorem 5, we then show that the WGA of downsampling strictly decreases in π_0 by examining the derivative.
3. Finally we note that by Theorem 1, upweighting must learn the same model as downsampling

We present the two lemmas below and use them to complete the proof.

Lemma 5 *Let $\ell(\hat{y},y)=\|y-\hat{y}\|_2^2$, $\hat{y}\in\mathbb{R}$, $y\in\mathcal{Y}$. Under Assumptions A1 to A4,*

$$WGA(\theta^*_{SRM}) = \Phi\left(\frac{\|\Delta_D\|^2-\tilde{c}_{\pi_0}\|\Delta_C\|^2}{2\sqrt{\|\Delta_D\|^2+\tilde{c}^2_{\pi_0}\|\Delta_C\|^2}}\right), \tag{10}$$

where $\tilde{c}_{\pi_0} := (1-4\pi_0)/(1+2\pi_0(1-2\pi_0)\|\Delta_C\|^2)$ and $\|v\| := \sqrt{v^T\Sigma^{-1}v}$.

Proof We begin by deriving the individual accuracy terms $A^{(y,d)}(\theta)$, $(y,d) \in \{0,1\} \times \{S,T\}$ as follows:

$$\begin{aligned}
A^{(1,d)}(\theta) &:= P\left(\mathbb{1}\left\{w^T X + b > 1/2\right\} = Y \mid Y = 1, D = d\right) \\
&= P\left(w^T X + b > 1/2 \mid Y = 1, D = d\right) \\
&= 1 - \Phi\left(\frac{1/2 - (w^T \mu^{(1,d)} + b)}{\sqrt{w^T \Sigma w}}\right) \\
&= \Phi\left(\frac{w^T \mu^{(1,d)} + b - 1/2}{\sqrt{w^T \Sigma w}}\right),
\end{aligned}$$

$$\begin{aligned}
A^{(0,d)}(\theta) &:= P\left(\mathbb{1}\left\{w^T X + b > 1/2\right\} = Y \mid Y = 0, D = d\right) \\
&= P\left(w^T X + b \leq 1/2 \mid Y = 0, D = d\right) \\
&= \Phi\left(\frac{1/2 - (w^T \mu^{(0,d)} + b)}{\sqrt{w^T \Sigma w}}\right).
\end{aligned}$$

We now derive the optimal model parameters for SRM for any $\pi_0 \leq 1/4$. In the case of SRM, $c(y,d) = 1$ for $(y,d) \in \{0,1\} \times \{S,T\}$, so the optimal solution to (2) is

$$w^*_{\text{SRM}} = \text{Var}(X)^{-1}\text{Cov}(X,Y), \quad \text{and} \quad b^*_{\text{SRM}} = \mathbb{E}[Y] - w^T \mathbb{E}[X]. \tag{11}$$

Note that

$$\begin{aligned}
\mathbb{E}[Y] &= \frac{1}{2}, \\
\mathbb{E}[X] &= \mathbb{E}\left[\mathbb{E}[X|D,Y]\right] \\
&= \sum_{(y,d)\in\{0,1\}\times\{S,T\}} [\mathbb{1}(d = T)(\mu^{(0,T)} - \mu^{(0,S)}) + (1-y)\mu^{(0,S)} + y\mu^{(1,S)}]\pi^{(y,d)} \\
&= \frac{1}{2}(\mu^{(0,T)} + \mu^{(1,S)}).
\end{aligned}$$

Let $\pi^{(d|y)} := P(D = d|Y = y)$ for $(y,d) \in \mathcal{Y} \times \mathcal{D}$, $\mu^{(y)} := \mathbb{E}[X|Y = y]$ for $y \in \mathcal{Y}$ and $\bar{\Delta} := \mu^{(1)} - \mu^{(0)}$. We compute $\text{Var}(X)$ as follows:

$$\begin{aligned}
\text{Var}(X) &= \mathbb{E}[\text{Var}(X|Y)] + \text{Var}(\mathbb{E}[X|Y]) && (12) \\
&= \mathbb{E}\left[\mathbb{E}[\text{Var}(X|Y,D)|Y] + \text{Var}(\mathbb{E}[X|Y,D]|Y)\right] + \text{Var}(\mathbb{E}[X|Y]) && (13) \\
&= \Sigma + \mathbb{E}[\text{Var}(\mathbb{E}[X|Y,D]|Y)] + \text{Var}(\mathbb{E}[X|Y]) && (14) \\
&= \Sigma + \mathbb{E}[\text{Var}(\mathbb{1}(D = T)(\mu^{(0,T)} - \mu^{(0,S)}) + (1-Y)\mu^{(0,S)} + Y\mu^{(1,S)}|Y)] && (15) \\
&\qquad + \text{Var}(\mathbb{E}[X|Y]) && (16) \\
&= \Sigma + \Delta_C\Delta_C^T\mathbb{E}[\text{Var}(\mathbb{1}(D = T)|Y)] + \text{Var}(Y(\mu^{(1)} - \mu^{(0)}) + \mu^{(0)}) && (17) \\
&= \Sigma + \Delta_C\Delta_C^T\mathbb{E}[\text{Var}(D|Y)] + \bar{\Delta}\bar{\Delta}^T\text{Var}(Y) && (18) \\
&= \Sigma + \Delta_C\Delta_C^T\mathbb{E}[Y\pi^{(T|1)}\pi^{(S|1)} + (1-Y)\pi^{(T|0)}\pi^{(S|0)}] + \bar{\Delta}\bar{\Delta}^T\pi^{(1)}\pi^{(0)} && (19) \\
&= \Sigma + 2\pi_0(1-2\pi_0)\Delta_C\Delta_C^T + \frac{1}{4}\bar{\Delta}\bar{\Delta}^T. && (20)
\end{aligned}$$

Next, we compute $\mathrm{Cov}(X, Y)$ as follows:

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathbb{E}[\mathrm{Cov}(X, Y|Y)] + \mathrm{Cov}(\mathbb{E}[X|Y], \mathbb{E}[Y|Y]) && (21)\\
&= \mathrm{Cov}(\mathbb{E}[X|Y], Y) && (22)\\
&= \mathrm{Cov}(\mu^{(0)} + Y\bar{\Delta}, Y) && (23)\\
&= \mathrm{Cov}(Y\bar{\Delta}, Y) && (24)\\
&= \bar{\Delta}\mathrm{Var}(Y) && (25)\\
&= \pi^{(1)}\pi^{(0)}\bar{\Delta} && (26)\\
&= \frac{1}{4}\bar{\Delta}. && (27)
\end{aligned}$$

In order to write (20) and (27) only in terms of Δ_C and Δ_D and to see the effect of π_0, in the following lemma we provide the relationship between $\bar{\Delta}$, Δ_C and Δ_D.

Lemma 6 *Let $\bar{\Delta} \coloneqq \mu^{(1)} - \mu^{(0)}$. Then*

$$\Delta_D - \bar{\Delta} = (1 - 4\pi_0)\Delta_C$$

Proof We first note that

$$\mu^{(1)} = 2\pi_0(\mu^{(1,S)} - \mu^{(1,T)}) + \mu^{(1,T)} = 2\pi_0\Delta_C + \mu^{(1,T)}.$$

Similarly,

$$\mu^{(0)} = 2\pi_0(\mu^{(0,T)} - \mu^{(0,S)}) + \mu^{(0,S)} = -2\pi_0\Delta_C + \mu^{(0,S)}.$$

Combining these with the definitions of Δ_D and $\bar{\Delta}$, we get

$$\begin{aligned}
\Delta_D - \bar{\Delta} &= \Delta_D - (\mu^{(1)} - \mu^{(0)})\\
&= \mu^{(1,T)} - \mu^{(0,T)} - 2\pi_0\Delta_C - \mu^{(1,T)} - 2\pi_0\Delta_C + \mu^{(0,S)}\\
&= \mu^{(0,S)} - \mu^{(0,T)} - 4\pi_0\Delta_C\\
&= (1 - 4\pi_0)\Delta_C.
\end{aligned}$$

■

From (20), (27) and Theorem 6, we then obtain

$$w^*_{\mathrm{SRM}} = \frac{1}{4}\left(\Sigma + 2\pi_0(1 - 2\pi_0)\Delta_C\Delta_C^T + \frac{1}{4}\bar{\Delta}\bar{\Delta}^T\right)^{-1}\bar{\Delta}, \tag{28}$$

where $\bar{\Delta} \coloneqq \mu^{(1)} - \mu^{(0)} = \Delta_D - (1 - 4\pi_0)\Delta_C$, and

$$b^*_{\mathrm{SRM}} = \frac{1}{2} - \frac{1}{2}(w^*_{\mathrm{SRM}})^T(\mu^{(0,T)} + \mu^{(1,S)}). \tag{29}$$

Therefore, for $d \in \{S, T\}$,

$$A^{(1,d)}(\theta^*_{\mathrm{SRM}}) = \Phi\left(\frac{(w^*_{\mathrm{SRM}})^T\left(\mu^{(1,d)} - \frac{1}{2}(\mu^{(0,T)} - \mu^{(1,S)})\right)}{\sqrt{(w^*_{\mathrm{SRM}})^T\Sigma w^*_{\mathrm{SRM}}}}\right), \tag{30}$$

$$A^{(0,d)}(\theta^*_{\mathrm{SRM}}) = \Phi\left(\frac{-(w^*_{\mathrm{SRM}})^T\left(\mu^{(0,d)} - \frac{1}{2}(\mu^{(0,T)} - \mu^{(1,S)})\right)}{\sqrt{(w^*_{\mathrm{SRM}})^T\Sigma w^*_{\mathrm{SRM}}}}\right). \tag{31}$$

We can simplify the expressions in (30) and (31) by using the following relations:

$$\begin{aligned}
&\mu^{(0,R)} - \frac{1}{2}(\mu^{(0,T)} - \mu^{(1,S)}) \\
&\qquad = \frac{1}{2}(\mu^{(0,T)} - \mu^{(1,S)}) = \frac{1}{2}(\mu^{(0,T)} - \mu^{(1,T)} + \mu^{(1,T)} - \mu^{(1,S)}) = -\frac{1}{2}(\Delta_C + \Delta_D), \\
&\mu^{(0,S)} - \frac{1}{2}(\mu^{(0,T)} - \mu^{(1,S)}) \\
&\qquad = \frac{1}{2}(\mu^{(0,S)} - \mu^{(0,T)}) + \frac{1}{2}(\mu^{(0,S)} - \mu^{(1,S)}) = \frac{1}{2}(\Delta_C - \Delta_D), \\
&\mu^{(1,S)} - \frac{1}{2}(\mu^{(0,T)} - \mu^{(1,S)}) \\
&\qquad = \frac{1}{2}(\mu^{(1,S)} - \mu^{(0,T)}) = \frac{1}{2}(\mu^{(1,S)} - \mu^{(1,T)} + \mu^{(1,T)} - \mu^{(0,T)}) = \frac{1}{2}(\Delta_C + \Delta_D), \\
&\mu^{(1,T)} - \frac{1}{2}(\mu^{(0,T)} - \mu^{(1,S)}) \\
&\qquad = \frac{1}{2}(\mu^{(1,T)} - \mu^{(0,T)}) + \frac{1}{2}(\mu^{(1,T)} - \mu^{(1,S)}) = \frac{1}{2}(\Delta_D - \Delta_C).
\end{aligned}$$

Plugging these into (30) and (31) for each group $(y, d) \in \{0, 1\} \times \{S, T\}$ yields

$$\begin{aligned}
A^{(0,T)}(\theta^*_{\text{SRM}}) = A^{(1,S)}(\theta^*_{\text{SRM}}) &= \Phi\left(\frac{\frac{1}{2}(w^*_{\text{SRM}})^T(\Delta_C + \Delta_D)}{\sqrt{(w^*_{\text{SRM}})^T \Sigma w^*_{\text{SRM}}}}\right), \\
A^{(0,S)}(\theta^*_{\text{SRM}}) = A^{(1,T)}(\theta^*_{\text{SRM}}) &= \Phi\left(\frac{\frac{1}{2}(w^*_{\text{SRM}})^T(\Delta_D - \Delta_C)}{\sqrt{(w^*_{\text{SRM}})^T \Sigma w^*_{\text{SRM}}}}\right).
\end{aligned}$$

Thus,

$$\text{WGA}(\theta^*_{\text{SRM}}) = \min\left\{\Phi\left(\frac{\frac{1}{2}(w^*_{\text{SRM}})^T(\Delta_C + \Delta_D)}{\sqrt{(w^*_{\text{SRM}})^T \Sigma w^*_{\text{SRM}}}}\right), \Phi\left(\frac{\frac{1}{2}(w^*_{\text{SRM}})^T(\Delta_D - \Delta_C)}{\sqrt{(w^*_{\text{SRM}})^T \Sigma w^*_{\text{SRM}}}}\right)\right\} \tag{32}$$

In order to rewrite (28) to be able to simplify (32), we will use the following lemma.

Lemma 7 *Let $A \in \mathbb{R}^{m \times m}$ be symmetric positive definite (SPD) and $u, v \in \mathbb{R}^m$. Then*

$$(A + vv^T + uu^T)^{-1}u = c_u\left(A^{-1}u - c_v A^{-1}v\right)$$

with

$$c_u := \frac{1}{1 + u^T B^{-1} u} \quad \textit{and} \quad c_v := \frac{v^T A^{-1} u}{1 + v^T A^{-1} v}.$$

Proof Let $B \coloneqq A + vv^T$. Then

$$
\begin{aligned}
(A + vv^T + uu^T)^{-1}u &= (B + uu^T)^{-1}u \\
&= \left(B^{-1} - \frac{B^{-1}uu^TB^{-1}}{1 + u^TB^{-1}u}\right)u \qquad \text{(Sherman-Morrison formula)} \\
&= B^{-1}u - \frac{u^TB^{-1}u}{1 + u^TB^{-1}u}B^{-1}u \\
&= c_uB^{-1}u \\
&= c_u\left(A^{-1} - \frac{A^{-1}vv^TA^{-1}}{1 + v^TA^{-1}v}\right)u \qquad \text{(Sherman-Morrison formula).}
\end{aligned}
$$

The assumption that A is SPD guarantees that A^{-1} exists, B is SPD, and c_u and c_v are well-defined. ■

Applying Theorem 7 to (28) with $A = A_{\text{SRM}} \coloneqq \Sigma$, $u = \bar{\Delta}/2$ and $v = \sqrt{\beta}\Delta_C$, where $\beta \coloneqq 2\pi_0(1 - 2\pi_0)$, yields

$$
\begin{aligned}
w^*_{\text{SRM}} &= \gamma_{\text{SRM}}\left(\Sigma^{-1}\bar{\Delta} - \frac{\beta\Delta_C^T\Sigma^{-1}\bar{\Delta}}{1 + \beta\Delta_C^T\Sigma^{-1}\Delta_C}\Sigma^{-1}\Delta_C\right) \\
&= \gamma_{\text{SRM}}\left(\Sigma^{-1}\Delta_D - \frac{1 - 4\pi_0 + \beta\Delta_C^T\Sigma^{-1}\Delta_D}{1 + \beta\Delta_C^T\Sigma^{-1}\Delta_C}\Sigma^{-1}\Delta_C\right) \qquad (\text{using } \bar{\Delta} = \Delta_D - (1 - 4\pi_0)\Delta_C) \\
&= \gamma_{\text{SRM}}\left(\Sigma^{-1}\Delta_D - c_{\pi_0}\Sigma^{-1}\Delta_C\right)
\end{aligned}
$$

with

$$
\gamma_{\text{SRM}} \coloneqq \frac{1}{4 + \bar{\Delta}^TA_{\text{SRM}}^{-1}\bar{\Delta}} \quad \text{and} \quad c_{\pi_0} \coloneqq \frac{1 - 4\pi_0 + \beta\Delta_C^T\Sigma^{-1}\Delta_D}{1 + \beta\Delta_C^T\Sigma^{-1}\Delta_C}.
$$

Let $\|v\| \coloneqq \sqrt{v^T\Sigma^{-1}v}$ be the norm induced by the Σ^{-1}–inner product. Then

$$
\begin{aligned}
\text{WGA}(\theta^*_{\text{SRM}}) = \min\Bigg\{&\Phi\left(\frac{(1 - c_{\pi_0})\Delta_C^T\Sigma^{-1}\Delta_D + \|\Delta_D\|^2 - c_{\pi_0}\|\Delta_C\|^2}{2\|\Delta_D - c_{\pi_0}\Delta_C\|}\right), \\
&\Phi\left(\frac{-(c_{\pi_0} + 1)\Delta_C^T\Sigma^{-1}\Delta_D + \|\Delta_D\|^2 + c_{\pi_0}\|\Delta_C\|^2}{2\|\Delta_D - c_{\pi_0}\Delta_C\|}\right)\Bigg\}
\end{aligned}
$$

Under Assumption A4, $c_{\pi_0} = \tilde{c}_{\pi_0} \coloneqq (1 - 4\pi_0)/(1 + 2\pi_0(1 - 2\pi_0)\|\Delta_C\|^2)$ and

$$
\text{WGA}(\theta^*_{\text{SRM}}) = \min\left\{\Phi\left(\frac{\|\Delta_D\|^2 - \tilde{c}_{\pi_0}\|\Delta_C\|^2}{2\sqrt{\|\Delta_D\|^2 + \tilde{c}_{\pi_0}^2\|\Delta_C\|^2}}\right), \Phi\left(\frac{\|\Delta_D\|^2 + \tilde{c}_{\pi_0}\|\Delta_C\|^2}{2\sqrt{\|\Delta_D\|^2 + \tilde{c}_{\pi_0}^2\|\Delta_C\|^2}}\right)\right\}, \tag{33}
$$

where the first term is the accuracy of the minority groups and the second is that of the majority groups. In order to be able to compare the WGA of SRM with the WGA of DS, we show that under Assumption A4 the WGA of SRM is given by the majority accuracy term in (33). Since $\tilde{c}_{\pi_0} \geq 0$ for $\pi_0 \leq 1/4$, we have that

$$
\frac{\|\Delta_D\|^2 - \tilde{c}_{\pi_0}\|\Delta_C\|^2}{2\sqrt{\|\Delta_D\|^2 + \tilde{c}_{\pi_0}^2\|\Delta_C\|^2}} \leq \frac{\|\Delta_D\|^2 + \tilde{c}_{\pi_0}\|\Delta_C\|^2}{2\sqrt{\|\Delta_D\|^2 + \tilde{c}_{\pi_0}^2\|\Delta_C\|^2}} \quad \Leftrightarrow \quad \tilde{c}_{\pi_0}\|\Delta_C\|^2 \geq 0,
$$

which is satisfied for all $\pi_0 \leq 1/4$ with equality at $\pi_0 = 1/4$. Since Φ is increasing, we obtain

$$\mathrm{WGA}(\theta^*_{\mathrm{SRM}}) = \Phi\left(\frac{\|\Delta_D\|^2 - \tilde{c}_{\pi_0}\|\Delta_C\|^2}{2\sqrt{\|\Delta_D\|^2 + \tilde{c}^2_{\pi_0}\|\Delta_C\|^2}}\right). \tag{34}$$

■

Since DS is a special case of SRM with $\pi_0 = 1/4$, we only need to examine (10) as a function of π_0. Still under Assumption A4, we take the following derivative:

$$\frac{\partial}{\partial \pi_0}\frac{\|\Delta_D\|^2 - \tilde{c}_{\pi_0}\|\Delta_C\|^2}{2\sqrt{\|\Delta_D\|^2 + \tilde{c}^2_{\pi_0}\|\Delta_C\|^2}} = \frac{-\|\Delta_D\|^2\|\Delta_C\|^2\left(\tilde{c}_{\pi_0}+1\right)}{2\left(\|\Delta_D\|^2 + \tilde{c}^2_{\pi_0}\|\Delta_C\|^2\right)^{3/2}} \times \frac{-2(16\pi_0^2 - 8\pi_0 + 3)\|\Delta_C\|}{(1 + 2\pi_0(1-2\pi_0)\|\Delta_C\|^2)^2},$$

which is strictly positive for $\pi_0 \leq 1/4$. Thus, the WGA of SRM is strictly increasing as a function of π_0, i.e., for $\pi_0 \leq 1/4$,

$$\mathrm{WGA}(\theta^*_{\mathrm{SRM}}) \leq \mathrm{WGA}(\theta^*_{\mathrm{DS}}) \overset{(i)}{=} \Phi\left(\|\Delta_D\|/2\right), \tag{35}$$

with equality when $\pi_0 = 1/4$ and where (i) follows since $\tilde{c}_{\pi_0} = 0$ when $\pi_0 = 1/4$. Additionally, by Theorem 1,

$$\mathrm{WGA}(\theta^*_{\mathrm{UW}}) = \mathrm{WGA}(\theta^*_{\mathrm{DS}}) = \Phi\left(\|\Delta_D\|/2\right). \tag{36}$$

Appendix C. Proof of Theorem 4

Our proof can be outlined as involving four steps; these steps rely on Theorem 1 and include two new lemmas. We enumerate the steps below:

1. We first show in Theorem 8 that SRM is agnostic to domain label noise.
2. In Theorem 9, we show that the model learned after downsampling with noisy domain labels is equivalent to a vanilla SRM model learned under no domain noise with group prior
$$\frac{(1-p)\pi_0}{4\pi_0^{(p)}} + \frac{p\pi_0}{4\big(1/2 - \pi_0^{(p)}\big)}.$$
3. Using the WGA for SRM derived in Theorem 5 for clean data and the group prior derived in Theorem 9, we then show that the WGA of downsampling strictly decreases in p by examining the derivative.
4. Finally we note that by Theorem 1, upweighting must learn the same model as downsampling

We present the two lemmas below and use them to complete the proof.

Lemma 8 *SRM with no data augmentation is agnostic to the domain label noise p, i.e., the model learned by SRM in (2) in the setting of domain label noise is the same as that learned in the setting of clean domain labels (no noise).*

Proof Since SRM with no data augmentation does not use domain label information when learning a model, the model will remain unchanged under domain label noise. ■

Lemma 9 *The model learned in* (2) *after downsampling according to noisy domain labels using the noisy minority prior* $\pi_0^{(p)} := (1-p)\pi_0 + p(1/2 - \pi_0)$ *for* $p \in [0, 1/2]$ *is equivalent to learning the model with clean domain labels (no noise) and using the minority prior*

$$\pi_{DS}^{(p)} := \frac{(1-p)\pi_0}{4\pi_0^{(p)}} + \frac{p\pi_0}{4(1/2 - \pi_0^{(p)})}. \tag{37}$$

Proof We note that the model learned after downsampling is agnostic to domain labels, so only the true proportion of each group, not the noisy proportion, determines the model weights. We derive the equivalent *clean* prior. We do so by examining how true minority samples are affected by DS on the data with noisy domain labels. When DS is performed on the data with domain label noise, the true minority samples that are kept can be categorized as (i) those that are still minority samples in the noisy data and (ii) a proportion of those that have become majority samples in the noisy data.

The first type of samples appear with probability

$$(1-p)\pi_0, \tag{38}$$

i.e., the proportion of true minority samples whose domain was not flipped. The second type of samples are kept with probability

$$p\pi_0 \left(\frac{\pi_0^{(p)}}{1/2 - \pi_0^{(p)}} \right), \tag{39}$$

where the factor dependent on $\pi_0^{(p)}$ is the factor by which the size of the noisy majority groups will be reduced to be the same size as the noisy minority groups.

Therefore, the unnormalized true minority prior can be written as

$$(1-p)\pi_0 + p\pi_0 \left(\frac{\pi_0^{(p)}}{1/2 - \pi_0^{(p)}} \right). \tag{40}$$

We can repeat the same analysis for the majority groups to obtain the unnormalized true majority prior as

$$p(1/2 - \pi_0) + (1-p)(1/2 - \pi_0) \left(\frac{\pi_0^{(p)}}{1/2 - \pi_0^{(p)}} \right). \tag{41}$$

In order for the true minority and true majority priors to sum to one over the four groups, we divide by the normalization factor $4\pi_0^{(p)}$, so our final minority prior is given by

$$\frac{(1-p)\pi_0 + p\pi_0 \left(\frac{\pi_0^{(p)}}{1/2 - \pi_0^{(p)}} \right)}{4\pi_0^{(p)}} = \frac{(1-p)\pi_0}{4\pi_0^{(p)}} + \frac{p\pi_0}{4(1/2 - \pi_0^{(p)})}. \tag{42}$$

■

DS is usually a special case of SRM with $\pi_0 = 1/4$. However, since DS uses domain labels and therefore is not agnostic to noise, we need to use the prior derived in Theorem 9 to be able to analyze the effect of the noise p while still using the clean data parameters. Note that $\pi_{\text{DS}}^{(p)}$ defined in (37) decreases from $1/4$ to π_0 as the noise p increases from 0 to $1/2$. Since we can interpolate between $1/4$ to π_0 using p, we can therefore substitute π_0 in (10) with $\pi_{\text{DS}}^{(p)}$ and then examine the resulting expression as a function of p for any π_0. Using the WGA of SRM derived in Theorem 5, we take the following derivative:

$$\begin{aligned}
&\frac{\partial}{\partial p}\frac{\|\Delta_D\|^2 - \tilde{c}_{\pi_{\text{DS}}^{(p)}}\|\Delta_C\|^2}{2\sqrt{\|\Delta_D\|^2 + \tilde{c}^2_{\pi_{\text{DS}}^{(p)}}\|\Delta_C\|^2}} \\
&\quad = \frac{-\|\Delta_D\|^2\|\Delta_C\|^2\left(\tilde{c}_{\pi_{\text{DS}}^{(p)}}+1\right)}{2\left(\|\Delta_D\|^2 + \tilde{c}^2_{\pi_{\text{DS}}^{(p)}}\|\Delta_C\|^2\right)^{3/2}} \times \frac{-2\left(16\left(\pi_{\text{DS}}^{(p)}\right)^2 - 8\pi_{\text{DS}}^{(p)} + 3\right)\|\Delta_C\|}{\left(1 + 2\pi_{\text{DS}}^{(p)}\left(1 - 2\pi_{\text{DS}}^{(p)}\right)\|\Delta_C\|^2\right)^2} \\
&\qquad \times \frac{\pi_0(4\pi_0 - 1)(2\pi_0 - 1)(2p - 1)}{2\left(\pi_0(2-4p)+p\right)^2\left(\pi_0(4p-2) - p + 1\right)^2},
\end{aligned}$$

which is strictly negative for $p < 1/2$ and $\pi_0 < 1/4$. Therefore, for any $\pi_0 < 1/4$, the WGA of DS is strictly decreasing in p and recovers the WGA of SRM when $p = 1/2$ or when $\pi_0 = 1/4$. Thus, for $p \leq 1/2$ and $\pi_0 \leq 1/4$,

$$\mathrm{WGA}(\theta^*_{\text{SRM}}) \leq \mathrm{WGA}(\theta^{(p)}_{\text{DS}}) = \Phi\left(\frac{\|\Delta_D\|^2 - \tilde{c}_{\pi_{\text{DS}}^{(p)}}\|\Delta_C\|^2}{2\sqrt{\|\Delta_D\| + \tilde{c}_{\pi_{\text{DS}}^{(p)}}\|\Delta_C\|}}\right), \tag{43}$$

with equality when $p = 1/2$ or $\pi_0 = 1/4$. Additionally, by Theorem 1,

$$\mathrm{WGA}(\theta^{(p)}_{\text{UW}}) = \mathrm{WGA}(\theta^{(p)}_{\text{DS}}) = \Phi\left(\frac{\|\Delta_D\|^2 - \tilde{c}_{\pi_{\text{DS}}^{(p)}}\|\Delta_C\|^2}{2\sqrt{\|\Delta_D\| + \tilde{c}_{\pi_{\text{DS}}^{(p)}}\|\Delta_C\|}}\right). \tag{44}$$