Proof of Lemma 1.  By definition we have

\[
R_T(u) = \sum_{t=1}^{T} \langle w_t - u, g_t \rangle = \sum_{t=1}^{T} \langle w_t - u, g_t \rangle (v_t - \|u\|) + \|u\| \sum_{t=1}^{T} \langle z_t, g_t \rangle (v_t - \|u\|) \\
= R^V_T(\|u\|) + \|u\| R^Z_T(\frac{u}{\|u\|}).
\]

\[\square\]

B Details from section 3

Proof of Theorem 2.  For any fixed \( u \in W \), let \( r = \max \frac{u \|u\|}{\|u\|} \).  Note that by definition we have \( \frac{u \|u\|}{\|u\|} \in [0, 1] \) and \( \frac{ru \|u\|}{\|u\|} \in W \).  Therefore, similar to the proof of Lemma 1, we decompose the regret against \( u \) as:

\[
R_T(u) = \sum_{t=1}^{T} \langle w_t - u, g_t \rangle = \sum_{t=1}^{T} \langle z_t, g_t \rangle (v_t - \frac{\|u\|}{r}) + \frac{\|u\|}{r} \sum_{t=1}^{T} \langle z_t - \frac{ru \|u\|}{\|u\|}, g_t \rangle,
\]
which, by the guarantees of \( A_V \) and \( A_Z,^3 \) is bounded in expectation by

\[
\tilde{O}\left( \frac{\|u\|}{r} L \sqrt{T} + \frac{\|u\|}{r} dL \sqrt{T} \right).
\]

Finally noticing \( \frac{1}{c} \leq r \) by the definition of \( c \) finishes the proof.  \[\square\]

C Details from section 4

Proof of Lemma 3.  Denote by \( \hat{w}_t = v_t z_t \).  By Jensen’s inequality we have

\[
\sum_{t=1}^{T} E \left[ \ell_t(\hat{w}_t) - \ell_t(u) \right] = E \left[ \sum_{t=1}^{T} \ell_t^v(\hat{w}_t) - \ell_t(u) \right] + E \left[ \sum_{t=1}^{T} \ell_t^z(\hat{w}_t) - \ell_t^u(\hat{w}_t) \right] \\
\leq \sum_{t=1}^{T} E \left[ \ell_t^v(\hat{w}_t) - \ell_t(u) \right]. \tag{5}
\]

We now continue under the assumption that \( \ell_t \) is \( L \)-Lipschitz.  After completing the proof of the first equation of Lemma 3 we use the \( \beta \)-smoothness assumption to prove the second equation of Lemma 3.

\[\text{Note that the condition } |\langle z_t, g_t \rangle| \leq 1 \text{ in Algorithm 4 indeed holds in this case since } Z = W \subseteq B \text{ and } \|g_t\|_2 \leq L \text{ by the Lipschitzness condition.}\]
Using the $L$-Lipschitz assumption we proceed:

$$\sum_{t=1}^{T} \mathbb{E} [\ell_i^a(w_t) - \ell_t(u)] \leq \sum_{t=1}^{T} \mathbb{E} [\ell_i^a(w_t) - \ell_i^a(u)] + \sum_{t=1}^{T} \mathbb{E} [\ell_i^a(u) - \ell_t(u)]$$

$$\leq \sum_{t=1}^{T} \mathbb{E} [\ell_i^a(w_t) - \ell_i^a(u)] + \mathbb{E}[L|v_t||\delta_t|]$$

$$\leq \sum_{t=1}^{T} \mathbb{E} [\ell_i^a(w_t) - \ell_i^a(u)] + \mathbb{E}[\delta L|v_t]$$

$$= \sum_{t=1}^{T} \mathbb{E} [\ell_i^a(w_t) - \ell_i^a(u)] + \mathbb{E}[\delta L|v_t]$$

$$+ \sum_{t=1}^{T} \mathbb{E} [\ell_i^a(w_t) - \ell_i^a(\tilde{w}_t)]$$

$$\leq \sum_{t=1}^{T} \mathbb{E} [\ell_i^a(w_t) - \ell_i^a(u)] + 2 \mathbb{E}[\delta L|v_t]].$$

Now, by using the $L$-Lipschitz assumption once more we find that

$$\sum_{t=1}^{T} \mathbb{E} [\ell_i^a((1 - \alpha)u) - \ell_i^a(u)] \leq \alpha \|u\|_2 TL \quad (6)$$

By using equation (6), the convexity of $\ell_i^a$, and Lemma 2 we continue with:

$$\sum_{t=1}^{T} \mathbb{E} [\ell_i(w_t) - \ell_t(u)] \leq \sum_{t=1}^{T} \mathbb{E} [(\tilde{w}_t - (1 - \alpha)u, \tilde{g}_t)] + 2 \mathbb{E}[\delta L|v_t] + \alpha \|u\|_2 TL$$

$$= \sum_{t=1}^{T} \mathbb{E} \left[ (v_t - \frac{\|u\|}{r}) \langle z_t, \tilde{g}_t \rangle \right] + \mathbb{E} \left[ \frac{\|u\|}{r} \langle z_t - \tilde{u}, \tilde{g}_t \rangle \right]$$

$$+ \sum_{t=1}^{T} 2 \mathbb{E}[\delta L|v_t] + \alpha \|u\|_2 TL$$

$$= \sum_{t=1}^{T} \mathbb{E} \left[ \tilde{\ell}_t(v_t) - \tilde{\ell}_t \left( \frac{\|u\|}{r} \right) \right] + \sum_{t=1}^{T} \frac{\|u\|}{r} \mathbb{E} [\langle z_t - \tilde{u}, \tilde{g}_t \rangle]$$

$$+ 2T\delta L \frac{\|u\|}{r} + \alpha \|u\|_2 TL$$

where $\tilde{\ell}_t(v) = v \langle z_t, \tilde{g}_t \rangle + 2\delta L|v|$ as defined in Algorithm 5, $\tilde{u} = \frac{v}{\|v\|} (1 - \alpha)u$, and $r > 0$ is such that $\frac{r\tilde{u}}{\|r\tilde{u}\|} \in \mathcal{Z}$.

Finally, by using the convexity of $\tilde{\ell}_t$, plugging in the guarantee of $A_V$, and using Theorem 6 we conclude the proof of the first equation of Lemma 3:

$$\sum_{t=1}^{T} \mathbb{E} [\ell_i(w_t) - \ell_t(u)]$$

$$\leq 2T\delta L \frac{\|u\|}{r} + \mathbb{E} \left[ \sum_{t=1}^{T} (v_t - \frac{\|u\|}{r}) \partial \tilde{\ell}_t(v_t) \right] + \frac{\|u\|}{r} \mathbb{E} \left[ \sum_{t=1}^{T} \langle z_t - \tilde{u}, \tilde{g}_t \rangle \right] + \alpha \|u\|_2 TL$$

$$= \tilde{\mathcal{O}} \left( 1 + T\delta L \frac{\|u\|}{r} + \frac{\|u\|}{r} L_V \sqrt{T} + \frac{\|u\|dL}{\sqrt{\delta T}} + \alpha \|u\|_2 TL \right).$$
Next, we continue from equation (5) under the smoothness condition. Using the definition of smoothness we find
\[
\sum_{t=1}^{T} E [\ell_t^i(w_t) - \ell_t(u)] \leq \sum_{t=1}^{T} E [\ell_t^i(w_t) - \ell_t^i(u)] + \sum_{t=1}^{T} E [\ell_t^i(u) - \ell_t(u)] \\
\leq \sum_{t=1}^{T} E [\ell_t^i(w_t) - \ell_t^i(u)] + E \left[ \frac{1}{2} \beta |v_t|^2 \|s_t\|^2 \right] \\
= \sum_{t=1}^{T} E [\ell_t^i(w_t) - \ell_t^i(u)] + E \left[ \frac{1}{2} \delta^2 |v_t|^2 \beta \right] \\
= \sum_{t=1}^{T} E [\ell_t^i(w_t) - \ell_t^i(u)] + E \left[ \frac{1}{2} \delta^2 |v_t|^2 \beta \right] \\
+ \sum_{t=1}^{T} E [\ell_t^i(u) - \ell_t^i(w_t)] \\
\leq \sum_{t=1}^{T} E [\ell_t^i(w_t) - \ell_t^i(u)] + E \left[ \beta |v_t|^2 \right].
\]
Using equation (6), the convexity of \(\ell_t^i\), and Lemma 2 we continue with:
\[
\sum_{t=1}^{T} E [\ell_t(w) - \ell_t(u)] \\
\leq \sum_{t=1}^{T} E [(\tilde{w}_t - (1 - \alpha)u, \tilde{g}_t)] + E \left[ \beta \delta^2 |v_t|^2 \right] + \alpha \|u\|_2 TL \\
= \sum_{t=1}^{T} E \left[ \left( v_t - \frac{\|u\|}{r} \right) \langle z_t, \tilde{g}_t \rangle \right] + E \left[ \beta \delta^2 |v_t|^2 \right] + \sum_{t=1}^{T} \frac{\|u\|}{r} E [\langle z_t - \tilde{u}, \tilde{g}_t \rangle] + \alpha \|u\|_2 TL \\
= T \beta \delta^2 \left( \frac{\|u\|}{r} \right)^2 + \sum_{t=1}^{T} E \left[ \tilde{\ell}_t(v_t) - \tilde{\ell}_t \left( \frac{\|u\|}{r} \right) \right] + \sum_{t=1}^{T} \frac{\|u\|}{r} E [\langle z_t - \tilde{u}, \tilde{g}_t \rangle] + \alpha \|u\|_2 TL,
\]
where \(\tilde{\ell}_t(v) = v \langle z_t, \tilde{g}_t \rangle + \beta \delta^2 v^2\) as defined in Algorithm 5. Finally, by using the convexity of \(\tilde{\ell}_t\), plugging in the guarantee of \(\mathcal{A}_V\), and using Theorem 6 we conclude the proof:
\[
\sum_{t=1}^{T} E [\ell_t(w) - \ell_t(u)] \\
\leq T \beta \delta^2 \left( \frac{\|u\|}{r} \right)^2 + E \left[ \sum_{t=1}^{T} \left( v_t - \frac{\|u\|}{r} \right) \partial \ell_t(v_t) \right] + \frac{\|u\|}{r} E \left[ \sum_{t=1}^{T} \langle z_t - \tilde{u}, \tilde{g}_t \rangle \right] + \alpha \|u\|_2 TL \\
= \tilde{O} \left( 1 + T \beta \delta^2 \left( \frac{\|u\|}{r} \right)^2 + \frac{\|u\|}{r} L V \sqrt{T} + \frac{\|u\|}{r} \frac{dL}{\delta} \sqrt{T} + \alpha \|u\|_2 TL \right).
\]

\[\square\]

**Theorem 6.** Suppose that \(\ell_t(0) = 0\), that \(\ell_t\) is \(L\)-Lipschitz for all \(t\), and that \(\mathcal{Z} \subseteq B\). For \(u \in (1 - \alpha)\mathcal{Z}\), Online Gradient Descent on \((1 - \alpha)\mathcal{Z}\) with learning rate \(\eta = \sqrt{\frac{\delta^2}{(dL)^2 T}}\) satisfies
\[
E \left[ \sum_{t=1}^{T} \langle z_t - u, \tilde{g}_t \rangle \right] \leq 2 \frac{dL}{\delta} \sqrt{T}.
\]

**Proof.** The proof essentially follows from the work of Zinkevich [27], Flaxman et al. [13] and using the assumptions that \(\ell_t(0) = 0\) and that \(\ell_t\) is \(L\)-Lipschitz. We start by bounding the norm of the
where the first inequality is the Cauchy-Schwarz inequality and the second is due to equation (7)

By equation (8)

Proof of Theorem 4. Plugging in \( \alpha \)

C.1 Details of section 4.1

By using equation (7) and the regret bound of Online Gradient Descent [27] we find that

\[
\sum_{t=1}^{T} \langle z_t, \hat{g}_t \rangle - \min_{z \in (1-\alpha) \mathbb{Z}} \sum_{t=1}^{T} \langle z, \hat{g}_t \rangle \leq (1-\alpha) \frac{dL}{\delta} \leq \frac{2dL}{\delta},
\]

where the first inequality is the Cauchy-Schwarz inequality and the second is due to equation (7). Since \(|\partial \ell_t(v_t)| \leq |\langle z_t, \hat{g}_t \rangle| + 2\delta L = L_V\) we can use Lemma 3 to find

\[
E[R_T(u)] = \tilde{O} \left( \delta TL \|u\| + \|u\| \frac{dL}{\delta} \sqrt{T} + \alpha TL \|u\|_2 \right).
\]

Plugging in \( \alpha = 0 \) and \( \delta = \min \{1, \sqrt{dT^{-\frac{1}{4}}} \} \) completes the proof.

Proof of Theorem 4. By equation (8) \(|\langle z_t, \hat{g}_t \rangle| \leq \frac{2dL}{\delta}\). Since \( v_t \leq \frac{1}{\beta} \) we have that

\[
|\partial \ell_t(v_t)| \leq \frac{dL}{\delta} + 2|v_t| \beta \delta^2 \leq \frac{dL + 2\beta}{\delta} \leq \frac{\beta(dL + 2)}{\delta}
\]

If \( \|u\|_2 \leq \frac{1}{\beta^2} \) applying Lemma 3 with \( \alpha = 0 \) gives us

\[
E \left[ \sum_{t=1}^{T} \ell_t(w_t) - \ell_t(u) \right] = \tilde{O} \left( 1 + T \beta \delta^2 \|u\|_2^2 + \|u\| \frac{dL \beta}{\delta} \sqrt{T} \right).
\]

If \( \|u\|_2 > \frac{1}{\beta^2} \) then using the Lipschitz assumption on \( \ell_t \) and equation (9) with \( u = 0 \) gives us

\[
E \left[ \sum_{t=1}^{T} \ell_t(w_t) - \ell_t(u) \right] = E \left[ \sum_{t=1}^{T} \ell_t(w_t) - \ell_t(0) + \ell_t(0) - \ell_t(u) \right] = \tilde{O}(1 + \|u\|_2^2 LT) = \tilde{O}(1 + \|u\|_2^2 \delta^3 LT),
\]

where we used that \( \|u\|_2 \geq \frac{1}{\beta^2} \). Adding equations (9) and (10) gives

\[
E \left[ \sum_{t=1}^{T} \ell_t(w_t) - \ell_t(u) \right] = \tilde{O} \left( 1 + \|u\|_2^2 \delta^3 LT + T \beta \delta^2 \|u\|_2^2 + \|u\| \frac{\beta dL}{\delta} \sqrt{T} \right)
\]
Setting $\delta = \min\{1, (dL)^{1/3}T^{-1/6}\}$ gives us

$$
\mathbb{E} \left[ \sum_{t=1}^{T} \ell_t(w_t) - \ell_t(u) \right] = \tilde{O} \left( 1 + \max\{\|u\|^2, \|u\|\} \beta (dLT)^{2/3} + \max\{\|u\|^2, \|u\|\} dL^2 \sqrt{T} \right).
$$

C.2 Details of section 4.2

**Proof of Theorem 5.** First, to see that $z_t + \delta s_t \in \mathcal{W}$ recall that by assumption $\mathcal{W} \subseteq \mathcal{B}$. Since $\alpha = \delta$ we have that $z_t + \delta s_t \in (1 - \alpha)\mathcal{W} + \delta \mathcal{S} \subseteq (1 - \delta)\mathcal{W} + \delta \mathcal{W} = \mathcal{W}$. For any fixed $u \in \mathcal{W}$, let $r = \max_{u' \in \mathcal{W}} \frac{\|u\|}{\|u\|}$. Note that by definition we have $\frac{\|u\|}{r} \in [0, 1]$ and $\frac{ru}{\|u\|} \in \mathcal{W}$. By using equation (8) we can see that $|\partial \ell_t(v_t)| \leq \frac{dL}{\delta} + 2dL$. By definition, $\frac{1}{r} \leq c$. This implies that the regret of $A_V$ is $\tilde{O} \left( 1 + \frac{\|u\|}{\|u\|} dL \sqrt{T} \right)$. Applying Lemma 3 with the parameters above we find

$$
\mathbb{E} \left[ \sum_{t=1}^{T} \ell_t(w_t) - \ell_t(u) \right] = \tilde{O} \left( 1 + \frac{\|u\|}{\|u\|} \sqrt{T} + c \|u\|^4 \delta L \sqrt{T} + c \|u\|^2 \frac{dL}{\delta} \sqrt{T} \right).
$$

Finally, setting $\delta = \min\{1, \sqrt{dT}^{-1/4}\}$ completes the proof:

$$
\mathbb{E} \left[ \sum_{t=1}^{T} \ell_t(w_t) - \ell_t(u) \right] = \tilde{O} \left( 1 + (\|u\|^2 + c\|u\|) \sqrt{dT}^{3/4} + c\|u\|dL \sqrt{T} \right).
$$

$\Box$