Appendix

1 DPCG algorithm with Gaussian noise

For all $k \in [K]$, let $Y^{(k)} \sim \mathcal{N}(0, \sigma^2 I)$. In order to compute the ℓ_2-sensitivity of the gradient, We can write:

$$\begin{aligned}
\|\nabla f_D(x) - \nabla f_{D'}(x)\|_2 &= \sqrt{\sum_{i=1}^{|V|} |\nabla_i f_D(x) - \nabla_i f_{D'}(x)|^2} \\
&\leq \sqrt{\sum_{i=1}^{|V|} (2\Delta)^2} \\
&= 2\sqrt{|V|}\Delta.
\end{aligned}$$

The Gaussian mechanism combined with the basic composition theorem provide the following privacy guarantee for the DPCG algorithm with Gaussian noise.

Theorem 1. *(Dwork and Roth, 2014) Let $\epsilon \in (0, 1)$ be arbitrary. For $c^2 > 2\ln(1.25K/\delta)$, the DPCG algorithm under Gaussian noise with parameter $\sigma \geq \frac{2cK\sqrt{|V|}\Delta}{\epsilon}$ is (ϵ, δ)-differentially private.*

We now analyze the approximation guarantee in this setting. First, we remind the reader that the following holds using Lemma 1 of the paper:

$$\mathbb{E}[f(x)] \geq (1 - \frac{1}{e})f(x^*) - G_{\mathcal{D}} - \frac{LR^2}{2K},$$

where if $\mathcal{D} = \mathcal{N}(0, \sigma^2 I)$, we have $G_{\mathcal{D}} \leq 2\text{rank}(\mathcal{M})\mathbb{E}_{Y \sim \mathcal{N}(0,\sigma^2 I)}\|Y\|_\infty$ and $G_{\mathcal{D}} \leq \frac{2}{c_{\min}}\mathbb{E}_{Y \sim \mathcal{N}(0,\sigma^2 I)}\|Y\|_\infty$ for matroid and knapsack constraints respectively. For a $|V|$-dimensional Gaussian random vector $Y \sim \mathcal{N}(0, \sigma^2 I)$, we can write:

$$\begin{aligned}
&\mathbb{E}\|Y\|_\infty \leq \mathcal{O}(\sigma\sqrt{\ln(|V|)}), \\
&\mathbb{P}\big(\|Y\|_\infty - \sigma\sqrt{2\ln(|V|)} \leq 2\sigma\sqrt{\ln(K)}\big) \geq 1 - \frac{1}{K^2}.
\end{aligned}$$

Combining the above results and setting $\sigma = \frac{2cK\sqrt{|V|}\Delta}{\epsilon}$ for $c^2 > 2\ln(1.25K/\delta)$, the following holds for matroid and knapsack constraints respectively:

$$\begin{aligned}
\mathbb{E}[f(x)] &\geq (1 - \frac{1}{e})f(x^*) - \frac{LR^2}{2K} - \mathcal{O}(\frac{\text{rank}(\mathcal{M})K\sqrt{|V|\ln(|V|)\ln(K/\delta)}\Delta}{\epsilon}), \\
\mathbb{E}[f(x)] &\geq (1 - \frac{1}{e})f(x^*) - \frac{LR^2}{2K} - \mathcal{O}(\frac{K\sqrt{|V|\ln(|V|)\ln(K/\delta)}\Delta}{c_{\min}\epsilon}).
\end{aligned}$$

Also, with probability at least $1 - \frac{1}{K}$, we have:

$$\begin{aligned}
f(x) &\geq (1 - \frac{1}{e})f(x^*) - \frac{LR^2}{2K} - \mathcal{O}(\frac{\text{rank}(\mathcal{M})K\sqrt{|V|\ln(\max\{|V|, K\})\ln(K/\delta)}\Delta}{\epsilon}), \\
f(x) &\geq (1 - \frac{1}{e})f(x^*) - \frac{LR^2}{2K} - \mathcal{O}(\frac{K\sqrt{|V|\ln(\max\{|V|, K\})\ln(K/\delta)}\Delta}{c_{\min}\epsilon}).
\end{aligned}$$

Compared to the Laplace noise, the additive factor in the approximation guarantee using the Gaussian noise is smaller by an order of $\sqrt{|V|\ln(|V|)}$. However, this improved accuracy comes at the price of achieving (ϵ, δ)-differential privacy as opposed to ϵ-differential privacy using the Laplace noise.

Algorithm 1 FTRL template for Online Linear Optimization (Agarwal and Singh, 2017)

Input: Noise distribution $\mathcal{D}$, regularizer $g(x)$.
Initialize an empty binary tree B to compute differentially private estimates of $\sum_{s=1}^{t} \ell_s$.
Sample $n_0^1, \ldots, n_0^{\lceil \ln T \rceil}$ independently from $\mathcal{D}$.
$\tilde{L}_0 = \sum_{i=1}^{\lceil \ln T \rceil} n_0^i$.
for $t = 1, \ldots, T$ **do**
  Choose $x_t = \arg\min_{x \in \mathcal{X}} \big(\eta \langle x, \tilde{L}_{t-1} \rangle + g(x)\big)$.
  Observe ℓ_t and suffer a loss of $\langle \ell_t, x_t \rangle$.
  $(\tilde{L}_t, B) = \text{TBAP}(\ell_t, B, t, \mathcal{D}, T)$.
end for

2 Algorithm 1 of Agarwal and Singh (2017)

The DPMFW algorithm exploits Algorithm 1 of Agarwal and Singh (2017) for differentially private online linear optimization as a sub-routine. We explain this algorithm in more detail below. The algorithm is provided in Algorithm 1. Consider an online linear optimization problem over T rounds where at each round $t \in [T]$, the algorithm chooses an action $x_t \in \mathcal{X}$, $\mathcal{X}$ is the fixed domain set, and upon committing to this action, a loss vector ℓ_t is revealed and the algorithm incurs the loss $\langle \ell_t, x_t \rangle$. Algorithm 1 is identical to the well-known FTRL algorithm except the fact that instead of $\sum_{s=1}^{t-1} \ell_s$, a noisy partial sum of the loss vectors $\tilde{L}_{t-1}$ is used in the update rule. This noisy partial sum is obtained using the Tree Based Aggregation Protocol (TBAP) which was used in prior works as well (Dwork et al., 2010; Jain et al., 2012).

3 Missing proofs

3.1 Proof of Lemma 2

The upper bounds for R follow from $\|x\|_1 \leq \text{rank}(\mathcal{M})$, $\forall x \in P(\mathcal{M})$ and $\|x\|_1 \leq \frac{1}{c_{\min}}$, $\forall x \in \{x \in [0,1]^{|V|} : c^T x \leq 1\}$. Consider the (i,j)-th entry of the Hessian of f. Let $m_F = \max_{i \in V} F(\{i\})$. We can write:

$$\begin{aligned} |\nabla^2_{i,j} f(z)| &= |\mathbb{E}_{R \sim z}\big[F(R \cup \{i,j\}) - F(R \cup \{i\} \setminus \{j\}) - F(R \cup \{j\} \setminus \{i\}) + F(R \setminus \{i,j\})\big]| \\ &\leq \max\{F(\{i\}), F(\{j\})\} \\ &\leq m_F. \end{aligned}$$

Thus, for all $k \in [K]$ and $j \in V$, using the mean-value theorem, we have:

$$\begin{aligned} |\nabla_j f(x^{(k)} + \frac{1}{K} v_k) - \nabla_j f(x^{(k)})| &\leq \frac{1}{K} m_F |1^T v_k| \\ &= m_F \|\frac{1}{K} v_k\|_1. \end{aligned}$$

Therefore, we can conclude $\|\nabla f(x^{(k)} + \frac{1}{K} v_k) - \nabla f(x^{(k)})\|_\infty \leq m_F \|\frac{1}{K} v_k\|_1$ and thus, $L \leq m_F$.

3.2 Proof of Lemma 5

First, note that by definition, the function g is monotone DR-submodular. Thus, similar to the proof of Lemma 1 in the paper, we can write:

$$\begin{aligned}
g(x^{(k+1)}) - g(x^{(k)}) &\overset{(a)}{\geq} \frac{1}{K}\langle v_k, \nabla g(x^{(k)})\rangle - \frac{L}{2K^2}\|v_k\|_1^2 \\
&\overset{(b)}{\geq} \frac{1}{K}\langle x^*, \nabla g(x^{(k)})\rangle + \frac{1}{K}\langle x^* - v_k, Y^{(k)}\rangle - \frac{LR^2}{2K^2} \\
&\overset{(c)}{\geq} \frac{1}{K}\langle (x^* - x^{(k)}) \vee 0, \nabla g(x^{(k)})\rangle + \frac{1}{K}\langle x^* - v_k, Y^{(k)}\rangle - \frac{LR^2}{2K^2} \\
&\overset{(d)}{\geq} \frac{1}{K}\big(g(x^* \vee x^{(k)}) - g(x^{(k)})\big) + \frac{1}{K}\langle x^* - v_k, Y^{(k)}\rangle - \frac{LR^2}{2K^2} \\
&\overset{(e)}{\geq} \frac{1}{K}\big(g(x^*) - g(x^{(k)})\big) + \frac{1}{K}\langle x^* - v_k, Y^{(k)}\rangle - \frac{LR^2}{2K^2},
\end{aligned}$$

where (a) is due to L-smoothness of g, (b) follows from the update rule of the algorithm, (c) and (e) use the monotonocity of g and (d) exploits concavity of g along non-negative directions. Using the definition of $G_{\mathcal{D}}$, if we take expectation of both sides, and apply the inequality recursively for all $k \in [K]$, we obtain:

$$\mathbb{E}[g(x^{(K+1)})] - g(x^*) \geq (1 - \frac{1}{K})^K \big(\mathbb{E}[g(\underbrace{x^{(1)}}_{=0})] - g(x^*)\big) - G_{\mathcal{D}} - \frac{LR^2}{2K}.$$

Rearranging the terms and using the inequality $(1 - \frac{1}{K})^K \leq \frac{1}{e}$, we can write:

$$\mathbb{E}[g(x)] \geq (1 - \frac{1}{e})g(x^*) - G_{\mathcal{D}} - \frac{LR^2}{2K}.$$

Using the update rule of the algorithm, we have:

$$\ell^T x = \ell^T x^{(K+1)} = \frac{1}{K}\sum_{k=1}^{K} \ell^T v_k \geq \frac{1}{K}\sum_{k=1}^{K} \lambda = \lambda = \ell^T x^*,$$

where the inequality is due to the update rule of the algorithm for v_k. Also, using the definition of ℓ and DR-submodularity of f, the following holds:

$$\begin{aligned}
\ell^T x^* &= \sum_{i \in [|V|]} x_i^* \ell_i \\
&\geq (1 - \kappa_F) \sum_{i \in [|V|]} x_i^* \nabla_i f(0) \\
&\geq (1 - \kappa_F) f(x^*).
\end{aligned}$$

Putting the above inequalities together, we have:

$$\begin{aligned}
\mathbb{E}[f(x)] &= \mathbb{E}[g(x)] + \mathbb{E}[\ell^T x] \\
&\geq (1 - \frac{1}{e})g(x^*) + \ell^T x^* - G_{\mathcal{D}} - \frac{LR^2}{2K} \\
&\geq (1 - \frac{1}{e})f(x^*) - (1 - \frac{1}{e})\ell^T x^* + \ell^T x^* - G_{\mathcal{D}} - \frac{LR^2}{2K} \\
&= (1 - \frac{1}{e})f(x^*) + \frac{1}{e}\ell^T x^* - G_{\mathcal{D}} - \frac{LR^2}{2K} \\
&\geq (1 - \frac{1}{e})f(x^*) + \frac{1}{e}(1 - \kappa_F)f(x^*) - G_{\mathcal{D}} - \frac{LR^2}{2K} \\
&\geq (1 - \frac{\kappa_F}{e})f(x^*) - G_{\mathcal{D}} - \frac{LR^2}{2K}.
\end{aligned}$$

3.3 Proof of Theorem 4

Similar to the offline setting, assume that all utility functions $\{f_t\}_{t=1}^T$ are monotone DR-submodular and L-smooth with respect to the norm $\|\cdot\|_1$. We can write:

$$
\begin{aligned}
f_t(x_t^{(k+1)}) &\geq f_t(x_t^{(k)}) + \frac{1}{K}\langle v_t^{(k)}, \nabla f_t(x_t^{(k)})\rangle - \frac{L}{2K^2}\|v_t^{(k)}\|_1^2 \\
&\geq f_t(x_t^{(k)}) + \frac{1}{K}\langle v_t^{(k)} - x^*, \nabla f_t(x_t^{(k)})\rangle + \frac{1}{K}\langle x^*, \nabla f_t(x_t^{(k)})\rangle - \frac{LR^2}{2K^2}.
\end{aligned}
$$

We can use the DR-submodularity and monotonocity of the utility function f_t to write:

$$
\begin{aligned}
\langle x^*, \nabla f_t(x_t^{(k)})\rangle &\geq \langle (x^* - x_t^{(k)}) \vee 0, \nabla f_t(x_t^{(k)})\rangle \\
&\geq f_t(x^* \vee x_t^{(k)}) - f_t(x_t^{(k)}) \\
&\geq f_t(x^*) - f_t(x_t^{(k)}).
\end{aligned}
$$

Combining the above inequalities, we have:

$$
f_t(x_t^{(k+1)}) \geq f_t(x_t^{(k)}) + \frac{1}{K}f_t(x^*) - \frac{1}{K}f_t(x_t^{(k)}) + \frac{1}{K}\langle v_t^{(k)} - x^*, \nabla f_t(x_t^{(k)})\rangle - \frac{LR^2}{2K^2}.
$$

Rearranging the terms and taking sum over $t \in [T]$, we obtain:

$$
\sum_{t=1}^T \big(f_t(x^*) - f_t(x_t^{(k+1)})\big) \leq (1 - \frac{1}{K})\sum_{t=1}^T \big(f_t(x^*) - f_t(x_t^{(k)})\big) + \frac{1}{K}\sum_{t=1}^T \langle x^* - v_t^{(k)}, \nabla f_t(x_t^{(k)})\rangle + \frac{LR^2T}{2K^2}.
$$

Applying the above inequality recursively for all $k \in [K]$, we have:

$$
\sum_{t=1}^T \big(f_t(x^*) - f_t(\underbrace{x_t^{(K+1)}}_{=x_t})\big) \leq \underbrace{(1 - \frac{1}{K})^K}_{\leq 1/e}\sum_{t=1}^T \big(f_t(x^*) - f_t(\underbrace{x_t^{(1)}}_{=0})\big) + \frac{1}{K}\sum_{k=1}^K\sum_{t=1}^T \langle x^* - v_t^{(k)}, \nabla f_t(x_t^{(k)})\rangle + \frac{LR^2T}{2K}.
$$

Rearranging the terms, we can equivalently write:

$$
R_T \leq \frac{1}{K}\sum_{k=1}^K\sum_{t=1}^T \langle x^* - v_t^{(k)}, \nabla f_t(x_t^{(k)})\rangle + \frac{LR^2T}{2K}.
$$

Using Theorem 3.1 of Agarwal and Singh (2017) with the regularizer $g(x) = \sum_{i=1}^{|V|} x_i \ln(x_i)$, we have the following for all $k \in [K]$:

$$
\mathbb{E}\big[\sum_{t=1}^T \langle x^* - v_t^{(k)}, \nabla f_t(x_t^{(k)})\rangle\big] \leq \mathcal{O}(R\sqrt{T\ln|V|}) + W_{\mathcal{D}},
$$

where $W_{\mathcal{D}} := \mathbb{E}_{Z\sim\mathcal{D}'}\big[\max_{x\in P}\langle Z, x\rangle - \min_{x\in P}\langle Z, x\rangle\big]$ and $\mathcal{D}'$ is the distribution induced by the sum of $\lceil \ln T\rceil$ independent samples from $\mathcal{D} = \mathrm{Lap}^{|V|}(\lambda)$ or $\mathcal{D} = \mathcal{N}(0, \sigma^2 I)$. For matroid constraints, we can write:

$$
\begin{aligned}
\max_{x\in P}\langle Z, x\rangle - \min_{x\in P}\langle Z, x\rangle &\leq \|x\|_1\|Z\|_\infty + \|x\|_1\|Z\|_\infty \\
&= 2\|x\|_1\|Z\|_\infty \\
&\leq 2\mathrm{rank}(\mathcal{M})\lceil \ln T\rceil \|Y\|_\infty,
\end{aligned}
$$

where $Y \sim \mathcal{D}$. Therefore, $W_{\mathrm{Lap}} \leq 2\mathrm{rank}(\mathcal{M})\lceil \ln T\rceil \mathbb{E}\|Y\|_\infty$ holds. Similarly, we have $W_{\mathrm{Lap}} \leq \frac{2}{c_{\min}}\lceil \ln T\rceil \mathbb{E}\|Y\|_\infty$ for knapsack constraints. If $\mathcal{D} = \mathrm{Lap}^{|V|}(\lambda)$, we have:

$$
\mathbb{E}\|Y\|_\infty \leq \mathcal{O}(\lambda \ln(|V|)).
$$

If $\mathcal{D} = \mathcal{N}(0, \sigma^2 I)$, the following holds:

$$
\mathbb{E}\|Y\|_\infty \leq \mathcal{O}(\sigma\sqrt{\ln(|V|)}).
$$

Setting $\lambda = \frac{2m_F|V|\ln T\sqrt{2K\ln(1/\delta)}}{\epsilon}$ and using the result of Lemma 2, we have the following regret bound using the Laplace noise and under matroid and knapsack constraints respectively.

$$\mathbb{E}[R_T] \leq \mathcal{O}(\text{rank}(\mathcal{M})\sqrt{T\ln|V|}) + \frac{m_F(\text{rank}(\mathcal{M}))^2 T}{2K} + \mathcal{O}(\frac{\text{rank}(\mathcal{M})|V|\ln|V|\ln^2 T\sqrt{K\ln(1/\delta)}}{\epsilon}),$$
$$\mathbb{E}[R_T] \leq \mathcal{O}(\frac{\sqrt{T\ln|V|}}{c_{\min}}) + \frac{m_F T}{2c_{\min}^2 K} + \mathcal{O}(\frac{|V|\ln|V|\ln^2 T\sqrt{K\ln(1/\delta)}}{c_{\min}\epsilon}).$$

Also, we can use the advanced composition theorem to conclude that the algorithm is (ϵ, δ)-differentially private. Setting $\sigma^2 = \frac{8\beta^2 K\ln(1/\delta)}{\epsilon^2}\ln^2 T\ln(\frac{K\ln T}{\delta'})$, the regret bound using the Gaussian noise for matroid and knapsack constraints are as follows:

$$\mathbb{E}[R_T] \leq \mathcal{O}(\text{rank}(\mathcal{M})\sqrt{T\ln|V|}) + \frac{m_F(\text{rank}(\mathcal{M}))^2 T}{2K} + \mathcal{O}(\frac{\text{rank}(\mathcal{M})\sqrt{\ln|V|}\ln^2 T\sqrt{K\ln(1/\delta)\ln(\frac{K\ln T}{\delta'})}}{\epsilon}),$$
$$\mathbb{E}[R_T] \leq \mathcal{O}(\frac{\sqrt{T\ln|V|}}{c_{\min}}) + \frac{m_F T}{2c_{\min}^2 K} + \mathcal{O}(\frac{\sqrt{\ln|V|}\ln^2 T\sqrt{K\ln(1/\delta)\ln(\frac{K\ln T}{\delta'})}}{c_{\min}\epsilon}).$$

Similarly, the algorithm is $(\epsilon, \delta + \delta')$-differentially private using the Gaussian noise. Setting $K = \mathcal{O}(\sqrt{T})$ in the above inequalities, we obtain the regret bounds as stated.

References

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3–4):211–407, August 2014. ISSN 1551-305X. doi: 10.1561/0400000042. URL `https://doi.org/10.1561/0400000042`.

Naman Agarwal and Karan Singh. The price of differential privacy for online learning. volume 70 of *Proceedings of Machine Learning Research*, pages 32–40, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR. URL `http://proceedings.mlr.press/v70/agarwal17a.html`.

Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N. Rothblum. Differential privacy under continual observation. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*, STOC '10, page 715–724, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781450300506. doi: 10.1145/1806689.1806787. URL `https://doi.org/10.1145/1806689.1806787`.

Prateek Jain, Pravesh Kothari, and Abhradeep Thakurta. Differentially private online learning. volume 23 of *Proceedings of Machine Learning Research*, pages 24.1–24.34, Edinburgh, Scotland, 25–27 Jun 2012. JMLR Workshop and Conference Proceedings. URL `http://proceedings.mlr.press/v23/jain12.html`.