Supplementary Material

Online Submodular Resource Allocation with Applications to Rebalancing Shared Mobility Systems

Pier Giuseppe Sessa, Ilija Bogunovic, Andreas Krause, Maryam Kamgarpour (ICML 2021)

A. Supplementary material for Section 4

In this section, we provide supplementary material for Section 4. First, we present two well-known lemmas which show important properties of the RKSH regression techniques of Section 3.2. Then, we define the notions of average and worst-case game curvature for (generally) non-differentiable welfare functions and state their main properties. Finally, we use these results to prove Thm 1 and Thm 2.

A.1. Confidence lemma and bound on posterior standard deviations

The following main lemma from Srinivas et al. (2010); Abbasi-Yadkori (2013); Chowdhury & Gopalan (2017) shows that the posterior mean and standard deviation functions computed in (4) can be used to construct a confidence interval around the unknown welfare functions $\gamma^r(\cdot)$.

Lemma 1. *Assume γ^r is a member of a RKHS with kernel function k^r and such that $\|\gamma^r\|_{k^r} \leq B$. Consider the observation model* (2) *and the posterior mean and standard deviation estimates $\mu_t^r(\cdot)$ and $\sigma_t^r(\cdot)$ computed as in* (4) *with regularization parameter $\lambda \geq 1$. Then, for any $\delta \in (0,1)$, with probability at least $1-\delta$,*

$$|\mu_t^r(\mathbf{x}, z) - \gamma^r(\mathbf{x}, z)| \leq \beta_t^r \sigma_t^r(\mathbf{x}, z), \quad \forall (\mathbf{x}, z) \in \boldsymbol{\mathcal{X}} \times \mathcal{Z}, \quad \forall t \geq 1$$

where $\beta_t^r = B + \sigma\lambda^{-1/2}\sqrt{2(g_t^r + \log(1/\delta))}$ and g_t^r is the maximum information gain defined in (7)*.*

Hence, according to Lemma 1, the ucb_t^r's functions defined in (5) represent a valid upper confidence bound on the true welfare. The following second lemma from, e.g., (Chowdhury & Gopalan, 2017, Lemma 4), bounds the sum of posterior standard deviations evaluated at the points selected by D-SUBUCB.

Lemma 2. *Consider the setup of Lemma 1, let $\{z_t\}_{t=1}^T$ be the sequence of observed contexts and $\{\mathbf{x}_t\}_{t=1}^T$ be the allocations selected by* D-SUBUCB*. Then,*

$$\sum_{t=1}^{T} \sigma_t^r(\mathbf{x}_t, z_t) \leq \sqrt{4T\lambda g_T^r}, \tag{11}$$

where g_t^r is the maximum information gain defined in (7)*.*

A.2. Game curvatures, general definitions and properties

Def 6 (Average and worst-case game curvature)**.** *Consider a sequence of contexts $z_1, \dots, z_T$. We define* average game curvature *and* worst-case game curvature*, respectively the quantities:*

$$c_{avg}(\{z_t\}_{t=1}^T) := 1 - \inf_i \lim_{k\to 0+} \frac{\sum_{t=1}^T \gamma(2\mathbf{x}_{max}, z_t) - \gamma(2\mathbf{x}_{max} - k\mathbf{e}_i, z_t)}{\sum_{t=1}^T \gamma(k\mathbf{e}_i, z_t) - \gamma(\mathbf{0}, z_t)} \quad \in [0,1]\,,$$

$$c_{wc}(\{z_t\}_{t=1}^T) := 1 - \inf_{t,i} \lim_{k\to 0+} \frac{\gamma(2\mathbf{x}_{max}, z_t) - \gamma(2\mathbf{x}_{max} - k\mathbf{e}_i, z_t)}{\gamma(k\mathbf{e}_i, z_t) - \gamma(\mathbf{0}, z_t)} \quad \in [0,1]\,,$$

where $\mathbf{x}_{max} = x_{max}\mathbf{1}$.

Note that when $\gamma(\cdot, z)$ is continuously differentiable, these can be equivalently defined as

$$c_{\text{avg}}(\{z_t\}_{t=1}^T) = 1 - \inf_i \frac{\sum_{t=1}^T \left[\nabla\gamma(2\mathbf{x}_{\max}, z_t)\right]_i}{\sum_{t=1}^T \left[\nabla\gamma(\mathbf{0}, z_t)\right]_i}, \qquad c_{\text{wc}}(\{z_t\}_{t=1}^T) = 1 - \inf_{t,i} \frac{\left[\nabla\gamma(2\mathbf{x}_{\max}, z_t)\right]_i}{\left[\nabla\gamma(\mathbf{0}, z_t)\right]_i}\,.$$

When the game is time-invariant (i.e., $z_t = \bar{z}, \forall t$) both these notions coincide with the definition of game curvature of Sessa et al. (2019b, Definition 2). Instead, for general contexts' sequences, $c_{\text{avg}}(\{z_t\}_{t=1}^T)$ represents the curvature of the average

game function $\gamma_{\text{avg}}(\cdot) = \sum_{t=1}^{T} \gamma(\cdot, z_t)$, while $c_{\text{wc}}(\{z_t\}_{t=1}^{T})$ quantifies the worst-case curvature over the game rounds. The following lemma states their main properties which we use to prove Thm 1 and Thm 2.

Lemma 3 (Properties of game curvatures)**.** *Consider the average and worst-case game curvatures defined in Def 6. We can affirm the following:*

(i) For any sequence of contexts $\{z_t\}_{t=1}^{T}$*,*

$$c_{avg}(\{z_t\}_{t=1}^{T}) \leq c_{wc}(\{z_t\}_{t=1}^{T}),$$

(ii) For any sequence of contexts $\{z_t\}_{t=1}^{T}$*, allocations* $\{\mathbf{x}_t\}_{t=1}^{T}$ *with* $\mathbf{x}_t \in \boldsymbol{\mathcal{X}}$*, and allocation* $\mathbf{y} \in \boldsymbol{\mathcal{X}}$*,*

$$\sum_{t=1}^{T} \gamma(\mathbf{x}_t + \mathbf{y}, z_t) - \gamma(\mathbf{x}_t, z_t) \geq \big(1 - c_{avg}(\{z_t\}_{t=1}^{T})\big)\Big[\sum_{t=1}^{T} \gamma(\mathbf{y}, z_t) - \gamma(\mathbf{0}, z_t)\Big].$$

(iii) For any sequence of contexts $\{z_t\}_{t=1}^{T}$*, allocations* $\{\mathbf{x}_t, \mathbf{y}_t\}_{t=1}^{T}$ *with* $\mathbf{x}_t, \mathbf{y}_t \in \boldsymbol{\mathcal{X}}$*,*

$$\sum_{t=1}^{T} \gamma(\mathbf{x}_t + \mathbf{y}_t, z_t) - \gamma(\mathbf{x}_t, z_t) \geq \big(1 - c_{wc}(\{z_t\}_{t=1}^{T})\big)\Big[\sum_{t=1}^{T} \gamma(\mathbf{y}_t, z_t) - \gamma(\mathbf{0}, z_t)\Big].$$

Proof. (i) Property *(i)* can be proved by showing that

$$\frac{\sum_{t=1}^{T} \gamma(2\mathbf{x}_{\max}, z_t) - \gamma(2\mathbf{x}_{\max} - k\mathbf{e}_i, z_t)}{\sum_{t=1}^{T} \gamma(k\mathbf{e}_i, z_t) - \gamma(\mathbf{0}, z_t)} \geq \inf_t \frac{\gamma(2\mathbf{x}_{\max}, z_t) - \gamma(2\mathbf{x}_{\max} - k\mathbf{e}_i, z_t)}{\gamma(k\mathbf{e}_i, z_t) - \gamma(\mathbf{0}, z_t)}, \tag{12}$$

for any index i and scalar $k \geq 0$. For simplicity, define $a_t = \gamma(2\mathbf{x}_{\max}, z_t) - \gamma(2\mathbf{x}_{\max} - k\mathbf{e}_i, z_t)$ and $b_t = \gamma(k\mathbf{e}_i, z_t) - \gamma(\mathbf{0}, z_t)$. Note that $a_t, b_t \geq 0$ for all t, by monotonicity of $\gamma(\cdot, z_t)$. Let,

$$\bar{t} = \arg\inf_t \frac{a_t}{b_t}. \tag{13}$$

Then, the following condition follows directly from (13):

$$a_t \geq a_{\bar{t}} \cdot \frac{b_t}{b_{\bar{t}}} \quad \forall t.$$

Using the above condition, we can lower bound the left hand side of (12) as

$$\frac{\sum_{t=1}^{T} \gamma(2\mathbf{x}_{\max}, z_t) - \gamma(2\mathbf{x}_{\max} - k\mathbf{e}_i, z_t)}{\sum_{t=1}^{T} \gamma(k\mathbf{e}_i, z_t) - \gamma(\mathbf{0}, z_t)} = \frac{\sum_{t=1}^{T} a_t}{\sum_{t=1}^{T} b_t} \geq \frac{a_{\bar{t}} \cdot \sum_{t=1}^{T} b_t / b_{\bar{t}}}{\sum_{t=1}^{T} b_t} = \frac{a_{\bar{t}}}{b_{\bar{t}}} \quad = \inf_t \frac{\gamma(2\mathbf{x}_{\max}, z_t) - \gamma(2\mathbf{x}_{\max} - k\mathbf{e}_i, z_t)}{\gamma(k\mathbf{e}_i, z_t) - \gamma(\mathbf{0}, z_t)},$$

which proves (12). □

(ii) Let us define the average game welfare $\gamma_{\text{avg}}(\cdot) = \sum_{t=1}^{T} \gamma(\cdot, z_t)$. Then, note that the average game curvature $c_{\text{avg}}(\{z_t\}_{t=1}^{T})$ coincides with the curvature (Sessa et al., 2019a, Definition 2) of γ_{avg} with respect to the set $[0, 2x_{\max}]^{NR}$. Then,

$$\begin{aligned}\sum_{t=1}^{T} \gamma(\mathbf{x}_t + \mathbf{y}, z_t) - \gamma(\mathbf{x}_t, z_t) &\geq \sum_{t=1}^{T} \gamma(2\mathbf{x}_{\max}, z_t) - \gamma(2\mathbf{x}_{\max} - \mathbf{y}, z_t) \\ &= \gamma_{\text{avg}}(2\mathbf{x}_{\max}) - \gamma_{\text{avg}}(2\mathbf{x}_{\max} - \mathbf{y}) \geq \big(1 - c_{\text{avg}}(\{z_t\}_{t=1}^{T})\big)\Big[\gamma_{\text{avg}}(\mathbf{y}) - \gamma_{\text{avg}}(\mathbf{0})\Big],\end{aligned}$$

where the first inequality is by DR-submodularity of $\gamma(\cdot, z)$ in each context z_t, and the second one follows directly by (Sessa et al., 2019a, Proposition 3). □

(iii) Note that the worst-case curvature $c_{\text{wc}}(\{z_t\}_{t=1}^T)$ coincides with the largest curvature (as per Sessa et al., 2019a, Definition 2) among the curvatures of the functions $\{\gamma(\cdot, z_t), t = 1, \dots, T\}$ with respect to the set $[0, 2x_{\max}]^{NR}$. Therefore, property *(iii)* holds since:

$$\sum_{t=1}^{T} \gamma(\mathbf{y}_t + \mathbf{x}_t, z_t) - \gamma(\mathbf{x}_t, z_t) \geq \sum_{t=1}^{T} \big(1 - c_{\text{wc}}(\{z_t\}_{t=1}^T)\big)\Big[\gamma(\mathbf{y}_t, z_t) - \gamma(\mathbf{0}, z_t)\Big]$$
$$= \big(1 - c_{\text{wc}}(\{z_t\}_{t=1}^T)\big)\Big[\sum_{t=1}^{T} \gamma(\mathbf{y}_t, z_t) - \gamma(\mathbf{0}, z_t)\Big],$$

where we have applied (Sessa et al., 2019a, Proposition 3) to each function $\gamma(\cdot, z_t)$ and used the fact that $c_{\text{wc}}(\{z_t\}_{t=1}^T)$ is the largest among their curvatures. □

A.3. Proof of Thm 1 and Corollary 1

Thm 1. *Consider the setup of Section 2. When* D-SUBUCB *is run with TW design (rule* (6)*) and β_t^r's are set according to Lemma 1, with probability at least $1 - \delta$,*

$$\sum_{t=1}^{T} \gamma(\mathbf{x}_t, z_t) \geq \alpha \cdot \text{OPT} - N \sum_{t=1}^{T} \sum_{r=1}^{R} 2\beta_t^r \sigma_t^r(\mathbf{x}_t, z_t) - \sum_{i=1}^{N} R^i(T)\,,$$

with $\alpha = \max\Big\{1 - c_{avg}(\{z_t\}_{t=1}^T),\ \big(1 + c_{wc}(\{z_t\}_{t=1}^T)\big)^{-1}\Big\}$.

Note that in the case of time-invariant games (i.e., $z_t = \bar{z}, \forall t$), $c_{\text{avg}}\big(\{z_t\}_{t=1}^T\big) = c_{\text{wc}}(\{z_t\}_{t=1}^T) = c$ as stated in Appendix A.2 and the above approximation guarantee $\alpha = \max\{(1-c),\ (1+c)^{-1}\} = (1+c)^{-1}$ coincides with the guarantee by (Vetta, 2002; Sessa et al., 2019a). For general context sequences, however, α depends on both notions of curvature.

Proof. Let $\mathbf{x}_\star = \arg\max_{\mathbf{x} \in \boldsymbol{\mathcal{X}}} \sum_{t=1}^T \gamma(\mathbf{x}, z_t)$ be the optimal action in hindsight. Moreover, define $\mathbf{x}_\star^{1:i} = [x_\star^1, \dots, x_\star^i, 0, \dots, 0]$ with $\mathbf{x}_\star^{1:0} = \mathbf{0}$. For ease of notation, let $\sigma_t = \sum_{r=1}^R \beta_t^r \sigma_t^r(\mathbf{x}_t, z_t)$. To bound the performance of D-SUBUCB, we will condition on the event of Lemma 1 holding true. Then, with probability $1 - \delta$, we can lower bound the obtained cumulative welfare as

$$\sum_{t=1}^{T} \gamma(\mathbf{x}_t, z_t) \geq \sum_{t=1}^{T} \sum_{i=1}^{N} \gamma(\mathbf{x}_t, z_t) - \gamma(0, x_t^{-i}, z_t) \tag{14}$$

$$\text{(Lemma 1)} \quad \geq \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{r=1}^{R} \text{ucb}_t^r(\mathbf{x}_t, z_t) - \gamma(0, x_t^{-i}, z_t) - N \sum_{t=1}^{T} 2\sigma_t \tag{15}$$

$$\text{(Def. of Regret)} \quad \geq \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{r=1}^{R} \text{ucb}_t^r(x_\star^i, x_t^{-i}, z_t) - \gamma(0, x_t^{-i}, z_t) - N \sum_{t=1}^{T} 2\sigma_t - \sum_{i=1}^{N} R^i(T) \tag{16}$$

$$\text{(Lemma 1)} \quad \geq \sum_{t=1}^{T} \sum_{i=1}^{N} \gamma(x_\star^i, x_t^{-i}, z_t) - \gamma(0, x_t^{-i}, z_t) - N \sum_{t=1}^{T} 2\sigma_t - \sum_{i=1}^{N} R^i(T) \tag{17}$$

$$\text{(DR-submodularity)} \quad \geq \sum_{t=1}^{T} \sum_{i=1}^{N} \gamma(\mathbf{x}_t + \mathbf{x}_\star^{1:i}, z_t) - \gamma(\mathbf{x}_t + \mathbf{x}_\star^{1:i-1}, z_t) - N \sum_{t=1}^{T} 2\sigma_t - \sum_{i=1}^{N} R^i(T) \tag{18}$$

$$\text{(telescoping sum)} \quad = \sum_{t=1}^{T} \gamma(\mathbf{x}_\star + \mathbf{x}_t, z_t) - \gamma(\mathbf{x}_t, z_t) - N \sum_{t=1}^{T} 2\sigma_t - \sum_{i=1}^{N} R^i(T). \tag{19}$$

The first inequality simply follows applying DR-submodularity of $\gamma(\cdot, z_t)$ in each context z_t (by DR-submodularity, $\gamma(\mathbf{x_t}, z_t)$ is at least the sum of its marginal contributions, see, e.g., Sessa et al. (2019b, Proof of Fact 1)), while the second one follows from Lemma 1 and the definition of the upper confidence bound functions ucb_t^r's in (5). Inequality (16) follows from the definition of players' regret $R^i(T)$ (Def 2) when the reward functions are computed according to the TW design rule (6).

At this point, we can apply the properties of the game curvatures stated in Lemma 3. By applying property *(ii)* to the bound (19) we get:

$$\sum_{t=1}^{T}\gamma(\mathbf{x}_t,z_t)\geq\big(1-c_{\text{avg}}(\{z_t\}_{t=1}^{T})\underbrace{\sum_{t=1}^{T}\gamma(\mathbf{x}_\star,z_t)}_{\text{OPT}}-N\sum_{t=1}^{T}2\sigma_t-\sum_{i=1}^{N}R^i(T) \tag{20}$$

where we have used the assumption $\gamma(\mathbf{0},z)=0,\forall z$. At the same time, by property *(iii)* we also have:

$$\sum_{t=1}^{T}\gamma(\mathbf{x}_\star+\mathbf{x}_t,z_t)-\gamma(\mathbf{x}_\star,z_t)\geq\big(1-c_{\text{wc}}(\{z_t\}_{t=1}^{T})\big)\big[\sum_{t=1}^{T}\gamma(\mathbf{x}_t,z_t)-\gamma(\mathbf{0},z_t)\big].$$

Therefore, after rearranging the previous bound and applying it to (19), we can lower bound the cumulative welfare also as:

$$\begin{aligned}\sum_{t=1}^{T}\gamma(\mathbf{x}_t,z_t)&\geq\underbrace{\sum_{t=1}^{T}\gamma(\mathbf{x}_\star,z_t)}_{\text{OPT}}-c_{\text{wc}}(\{z_t\}_{t=1}^{T})\cdot\gamma(\mathbf{x}_t,z_t)-N\sum_{t=1}^{T}2\sigma_t-\sum_{i=1}^{N}R^i(T)\\&\geq\frac{1}{1+c_{\text{wc}}(\{z_t\}_{t=1}^{T})}\text{OPT}\;-N\sum_{t=1}^{T}2\sigma_t-\sum_{i=1}^{N}R^i(T)\end{aligned} \tag{21}$$

Hence, the theorem statement is obtained combining bounds (20) and (21). □

As outlined in Section 4.1, from Thm 1 we can obtain the following corollary.

Corollary 1. *Consider the setup of Section 2 and assume $|\mathcal{X}^i|=K$ for all i. Then, if* D-SUBUCB *is run with TW design, $\beta_t^r=B+\sigma\lambda^{-1/2}\sqrt{2(g_t^r+\log(2/\delta))}$ and* NO-REGRET *is* MWU *(Algorithm 1), with probability $1-\delta$,*

$$\sum_{t=1}^{T}\gamma(\mathbf{x}_t,z_t)\geq\alpha\cdot\text{OPT}-N\sum_{r=1}^{R}\mathcal{O}\big(g_T^r\sqrt{T}\big)-N\cdot\mathcal{O}\big(\sqrt{T\log K}+\sqrt{T\log(2/\delta)}\big)\,,$$

with $\alpha=\max\Big\{1-c_{avg}(\{z_t\}_{t=1}^{T}),\;\big(1+c_{wc}(\{z_t\}_{t=1}^{T})\big)^{-1}\Big\}$.

Proof. The corollary can be obtained by bounding individually the terms in the statement of Thm 1. First, Lemma 2 implies that

$$N\sum_{t=1}^{T}\sum_{r=1}^{R}2\beta_t^r\sigma_t^r(\mathbf{x}_t,z_t)\leq 2N\beta_T^r\sum_{t=1}^{T}\sum_{r=1}^{R}\sigma_t^r(\mathbf{x}_t,z_t)\leq N\sum_{r=1}^{R}\mathcal{O}\big(g_T^r\sqrt{T}\big)\,. \tag{22}$$

Second, the well-known result from, e.g., Cesa-Bianchi & Lugosi (2006, Section 4.2) shows that, with probability at least $1-\delta_1$, the regret of MWU (Algorithm 1) can be bounded as

$$R^i(T)\leq\mathcal{O}\big(\sqrt{T\log K}+\sqrt{T\log(1/\delta_1)}\big) \tag{23}$$

Finally, the specific choice of β_t^r implies that the event in the confidence Lemma A.1 holds true with probability at least $1-\delta/2$. Hence, by setting $\delta_1=\delta/2$ and using (22),(23), a standard probability union bound shows that with probability at least $1-\delta/2-\delta/2=1-\delta$ the cumulative welfare can be lower bounded as stated in Corollary 1. □

A.4. Proof of Theorem 2

Thm 2. *Consider the setup of Section 2 and assume the game is anonymous and $\mathcal{X}^i=\{0,x_{max}\}^R,\forall i$. When* D-SUBUCB *is run with ES design (rule* (8)*) and β_t^r's are set according to Lemma 1, with probability at least $1-\delta$,*

$$\sum_{t=1}^{T}\gamma(\mathbf{x}_t,z_t)\geq\alpha\cdot\text{OPT}-\sum_{t=1}^{T}\sum_{r=1}^{R}2\beta_t^r\sigma_t^r(\mathbf{x}_t,z_t)-\sum_{i=1}^{N}R^i(T)-N\sum_{t=1}^{T}\sum_{r=1}^{R}\epsilon^r(z_t),$$

with $\alpha=\max\Big\{1-c_{avg}(\{z_t\}_{t=1}^{T}),\;\big(1+c_{wc}(\{z_t\}_{t=1}^{T})\big)^{-1}\Big\}$.

Proof. To prove Theorem 2, we first establish the following Lemma which shows an important property of the ES design rule (8).

Lemma 4. *Assume $\mathcal{X}^i = \{0, x_{max}\}^R$ for $i = 1, \dots, N$ and the game is* anonymous *as defined in Section 4.2. Then, consider any player i, resource r, and any strategy $\mathbf{x} = (x^i, x^{-i}) \in \boldsymbol{\mathcal{X}}$ such that $x^i[r] > 0$ (i.e., player i selects resource r). For each context $z \in \mathcal{Z}$ it holds:*

$$\frac{1}{|(x^i, x^{-i})|_r}\gamma^r(x^i, x^{-i}, z) \geq \gamma^r(x^i, x^{-i}, z) - \gamma(0, x^{-i}, z) - \epsilon^r(z)\,, \tag{24}$$

where $\epsilon^r(z)$ is the weak-separability error defined in Def 4.

Proof. Without loss of generality assume $i = 1$, and that only players $\{1, \dots, P\}$ select resource R, so that $|(x^i, x^{-i})|_r = P$. Moreover, we define $[\mathbf{x}]_r \in \boldsymbol{\mathcal{X}}$ to be the modified version of $\mathbf{x}$ where all the entries corresponding to resources different from r are set to 0 (hence, $[\mathbf{x}]_r$ has only P nonzero entries). Recall also in Section 4.2 we have defined $[x^i]_{-r}$ to be the modified version of x^i where $x^i[r]$ is set to zero. For simplicity we also drop the dependence of γ^r and ϵ^r on context z. We have:

$$\begin{aligned}
\frac{1}{|(x^i, x^{-i})|_r}\gamma^r(x^i, x^{-i}) &= \frac{1}{P}\gamma^r(x^i, x^{-i}) \geq \frac{1}{P}\gamma^r([x^i, x^{-i}]_r) = \frac{1}{P}\gamma^r([x^1, \dots, x^P, 0, \dots, 0]_r) \\
&= \frac{1}{P}\Big[\gamma^r([x^1, 0, \dots, 0]_r) - \gamma^r(\mathbf{0}) + \\
&\qquad + \sum_{i=2}^{P} \gamma^r([x^1, \dots, x^i, 0, \dots, 0]_r) - \gamma^r([0, x^2, \dots, x^i, 0, \dots, 0]_r)\Big] \qquad (25) \\
&\geq \frac{1}{P}\sum_{i=1}^{P} \gamma^r([x^1, \dots, x^P, 0, \dots, 0]_r) - \gamma^r([0, x^2 \dots, x^P, 0, \dots, 0]_r) \qquad (26) \\
&= \gamma^r([x^i, x^{-i}]_r) - \gamma^r([0, x^{-i}]_r) \\
&\geq \gamma^r(x^i, x^{-i}) - \gamma^r([x^i]_{-r}, x^{-i}) \qquad (27) \\
&= \gamma^r(x^i, x^{-i}) - \gamma^r(0, x^{-i}) - \big(\gamma([x^i]_{-r}, x^{-i}) - \gamma^r(0, x^{-i})\big) \\
&\geq \gamma^r(x^i, x^{-i}) - \gamma^r(0, x^{-i}) - \big(\gamma^r([x^i]_{-r}, 0) - \gamma^r(\mathbf{0})\big) \qquad (28) \\
&\geq \gamma^r(x^i, x^{-i}) - \gamma^r(0, x^{-i}) - \epsilon^r.
\end{aligned}$$

The first inequality is due to monotonicity, while (25) is a telescoping sum because the game is *anonymous* and since $\gamma^r(\mathbf{0}) = 0$. Then, (26) is obtained applying DR-submodularity to each summation term. Inequalities (27) and (28) are again due to DR-submodularity, while the last inequality follows by Def 4. □

We are now ready to prove Thm 2. First, let us consider a generic round t and let $R_t \subset [R]$ be the set of resources selected by at least 1 player, i.e., $R_t = \{r : \exists i : x_t^i[r] > 0\}$. Then, it holds:

$$\gamma(\mathbf{x}_t, z_t) = \sum_{r=1}^{R} \gamma^r(\mathbf{x}_t, z_t) \geq \sum_{r \in R_t} \gamma^r(\mathbf{x}_t, z_t) = \sum_{i=1}^{N} \sum_{r : x_t^i[r] > 0} \frac{1}{|\mathbf{x}_t|_r} \cdot \gamma^r(\mathbf{x}_t, z_t), \tag{29}$$

where in the first inequality we have used the fact that $\gamma^r(\mathbf{x}, z) \geq 0$ for all $\mathbf{x} \in \boldsymbol{\mathcal{X}}$ and $z \in \mathcal{Z}$ (since $\gamma^r(\mathbf{0}, z) = 0$ and $\gamma^r(\cdot, z)$ is monotone). We can now use (29) to prove Thm 2. As in proof of Thm 1, we let $\mathbf{x}_\star = \arg\max_{\mathbf{x} \in \boldsymbol{\mathcal{X}}} \sum_{t=1}^{T} \gamma(\mathbf{x}, z_t)$ be the optimal action in hindsight, $\mathbf{x}_\star^{1:i} = [x_\star^1, \dots, x_\star^i, 0, \dots, 0]$ with $\mathbf{x}_\star^{1:0} = \mathbf{0}$ and $\sigma_t = \sum_{r=1}^{R} \beta_t^r \sigma_t^r(\mathbf{x}_t, z_t)$. Then, conditioning

on the event of Lemma 1, with probability $1-\delta$, the obtained cumulative welfare can be lower bounded as follows:

$$\sum_{t=1}^{T}\gamma(\mathbf{x}_t,z_t) \geq \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r:x_t^i[r]>0}\frac{1}{|\mathbf{x}_t|_r}\cdot\gamma^r(\mathbf{x}_t,z_t)$$

$$\text{(Lemma 1)} \quad \geq \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r:x_t^i[r]>0}\frac{1}{|\mathbf{x}_t|_r}\cdot\mathrm{ucb}_t^r(\mathbf{x}_t,z_t) - 2\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r:x_t^i[r]>0}\frac{1}{|\mathbf{x}_t|_r}\cdot\beta_t^r\sigma_t^r(\mathbf{x}_t,z_t) \tag{30}$$

$$\text{(Def. of Regret)} \quad \geq \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r:x_\star^i[r]>0}\frac{1}{|(x_\star^i,x_t^{-i})|_r}\cdot\mathrm{ucb}_t^r(x_\star^i,x_t^{-i},z_t) - \sum_{t=1}^{T}2\sigma_t - \sum_{i=1}^{N}R^i(T) \tag{31}$$

$$\text{(Lemma 1)} \quad \geq \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r:x_\star^i[r]>0}\frac{1}{|(x_\star^i,x_t^{-i})|_r}\cdot\gamma^r(x_\star^i,x_t^{-i},z_t) - \sum_{t=1}^{T}2\sigma_t - \sum_{i=1}^{N}R^i(T) \tag{32}$$

$$\text{(Lemma 4)} \quad \geq \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r:x_\star^i[r]>0}\gamma^r(x_\star^i,x_t^{-i},z_t) - \gamma(0,x_t^{-i},z_t) - \epsilon^r(z_t) - \sum_{t=1}^{T}2\sigma_t - \sum_{i=1}^{N}R^i(T) \tag{33}$$

$$= \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r=1}^{R}\gamma^r(x_\star^i,x_t^{-i},z_t) - \gamma^r(0,x_t^{-i},z_t) - \sum_{t=1}^{T}2\sigma_t - \sum_{i=1}^{N}R^i(T) \tag{34}$$

$$- \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r:x_\star^i[r]=0}\gamma^r(x_\star^i,x_t^{-i},z_t) - \gamma^r(0,x_t^{-i},z_t) - \sum_{r:x_\star^i[r]>0}\epsilon^r(z_t)$$

$$\geq \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r=1}^{R}\gamma^r(x_\star^i,x_t^{-i},z_t) - \gamma^r(0,x_t^{-i},z_t) - \sum_{t=1}^{T}2\sigma_t - \sum_{i=1}^{N}R^i(T) - \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r=1}^{R}\epsilon^r(z_t) \tag{35}$$

$$= \sum_{t=1}^{T}\sum_{i=1}^{N}\gamma(x_\star^i,x_t^{-i},z_t) - \gamma(0,x_t^{-i},z_t) - \sum_{t=1}^{T}2\sigma_t - \sum_{i=1}^{N}R^i(T) - N\sum_{t=1}^{T}\sum_{r=1}^{R}\epsilon^r(z_t). \tag{36}$$

Inequality (30) is due to Lemma 1 and the definition of ucb_t, while (31) follows from the definition of players' regret (Def 2) when the rewards are computed according to ES design rule (8). Then, (32) is again due to Lemma 1 and (33) is obtained applying Lemma 4 for each time t, player i, and resource r such that $x_\star^i[r]>0$. In (34) we have added and subtracted, the term $\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r:x_\star^i[r]=0}\gamma^r(x_\star^i,x_t^{-i},z_t)-\gamma^r(0,x_t^{-i},z_t)$. Then, (35) is obtained since for each t, i, and r such that $x_\star^i[r]=0$,

$$\gamma^r(x_\star^i,x_t^{-i},z_t) - \gamma^r(0,x_t^{-i},z_t) \geq \gamma^r(x_\star^i,0,z_t) - \gamma^r(\mathbf{0},z_t) = \gamma^r([x_\star^i]_{-r},0,z_t) - \gamma^r(\mathbf{0},z_t) \leq e^r(z_t), \tag{37}$$

where the first inequality is due to DR-submodularity, the equality since $x_\star^i[r]=0$, and the last inequality by definition of weak-separability errors (Def 4). Finally, (36) follows from the definition of γ. From (36), the statement of the theorem is obtained following the same proof steps of Proof of Thm 1 in Appendix A.3 to lower bound the term $\sum_{t=1}^{T}\sum_{i=1}^{N}\gamma(x_\star^i,x_t^{-i},z_t)-\gamma(0,x_t^{-i},z_t)$ (see Equation (17) and subsequent proof steps). □

B. Stronger benchmark: seeking optimal policies

In this section we extend the results obtained in Section 4 to the case where context z_t is observed *before* choosing allocation $\mathbf{x}_t$ and we compete with the stronger performance benchmark of the optimal contextual welfare OPT_c defined in (10). As outlined in Section 5, in this richer setting the D-SUBUCB algorithm computes allocations by simulating a contextual game among the players, where each player is equipped with an algorithm having sublinear *contextual regret* $R_c^i(T)$, as defined in Def 5. The following theorem bounds the performance of D-SUBUCB under TW and ES design, respectively.

Thm 3. *Consider the setup of Section 2 and assume context z_t is observed before choosing allocation $\mathbf{x}_t$. Then, when* D-SUBUCB *is run with β_t^r's set according to Lemma 1, with probability at least 1-δ,*

1) If game rewards are computed according to TW design (rule (6)*),*

$$\sum_{t=1}^{T}\gamma(\mathbf{x}_t,z_t)\geq\alpha\cdot\mathrm{OPT}_c-N\sum_{t=1}^{T}\sum_{r=1}^{R}2\beta_t^r\sigma_t^r(\mathbf{x}_t,z_t)-\sum_{i=1}^{N}R_c^i(T)\,,$$

2) If the game is anonymous, $\mathcal{X}^i=\{0,x_{max}\}^R$, and game rewards are computed according to ES design (rule (8)*),*

$$\sum_{t=1}^{T}\gamma(\mathbf{x}_t,z_t)\geq\alpha\cdot\mathrm{OPT}_c-\sum_{t=1}^{T}\sum_{r=1}^{R}2\beta_t^r\sigma_t^r(\mathbf{x}_t,z_t)-\sum_{i=1}^{N}R_c^i(T)-N\sum_{t=1}^{T}\sum_{r=1}^{R}\epsilon^r(z_t),$$

where $\alpha=\big(1+c_{wc}(\{z_t\}_{t=1}^T)\big)^{-1}$.

Proof. The proofs of *1)* and *2)* follow closely the proofs of Thm 1 and Thm 2, respectively, with minor important differences. Let $\boldsymbol{\pi}_\star=\arg\max_{\boldsymbol{\pi}:\mathcal{Z}\to\boldsymbol{\mathcal{X}}}\sum_{t=1}^T\gamma(\boldsymbol{\pi}(z_t),z_t)$ be the optimal policy in hindsight. Moreover, denote with $\pi_\star^i(\cdot)$ the optimal policy in hindsight concerning player i, i.e., $\boldsymbol{\pi}_\star(z)=[\pi_\star^1(z),\dots,\pi_\star^N(z)]$ for each z. We also define $\boldsymbol{\pi}_\star^{1:i}(z)=[\pi_\star^1(z),\dots,\pi_\star^i(z),0,\dots,0]$ with $\boldsymbol{\pi}_\star^{1:0}(z)=\mathbf{0}$, and $\sigma_t=\sum_{r=1}^R\beta_t^r\sigma_t^r(\mathbf{x}_t,z_t)$.

1) Let us fist consider the case of TW design. Following the same proof steps as in Proof of Thm 1 (Appendix A.3),

$$\begin{aligned}
\sum_{t=1}^{T}\gamma(\mathbf{x}_t,z_t)&\geq\sum_{t=1}^{T}\sum_{i=1}^{N}\gamma(\mathbf{x}_t,z_t)-\gamma(0,x_t^{-i},z_t)\\
\text{(Lemma 1)}\quad&\geq\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r=1}^{R}\mathrm{ucb}_t^r(\mathbf{x}_t,z_t)-\gamma(0,x_t^{-i},z_t)-N\sum_{t=1}^{T}2\sigma_t\\
\text{(Def. of Contextual Regret)}\quad&\geq\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r=1}^{R}\mathrm{ucb}_t^r(\pi_\star^i(z_t),x_t^{-i},z_t)-\gamma(0,x_t^{-i},z_t)-N\sum_{t=1}^{T}2\sigma_t-\sum_{i=1}^{N}R_c^i(T)\qquad(38)\\
\text{(Lemma 1)}\quad&\geq\sum_{t=1}^{T}\sum_{i=1}^{N}\gamma(\pi_\star^i(z_t),x_t^{-i},z_t)-\gamma(0,x_t^{-i},z_t)-N\sum_{t=1}^{T}2\sigma_t-\sum_{i=1}^{N}R_c^i(T)\qquad(39)\\
\text{(DR-submodularity)}\quad&\geq\sum_{t=1}^{T}\sum_{i=1}^{N}\gamma(\mathbf{x}_t+\boldsymbol{\pi}_\star^{1:i}(z_t),z_t)-\gamma(\mathbf{x}_t+\boldsymbol{\pi}_\star^{1:i-1}(z_t),z_t)-N\sum_{t=1}^{T}2\sigma_t-\sum_{i=1}^{N}R_c^i(T)\\
\text{(telescoping sum)}\quad&=\sum_{t=1}^{T}\gamma(\boldsymbol{\pi}_\star(z_t)+\mathbf{x}_t,z_t)-\gamma(\mathbf{x}_t,z_t)-N\sum_{t=1}^{T}2\sigma_t-\sum_{i=1}^{N}R_c^i(T),\qquad(40)
\end{aligned}$$

where in (38) we have used the definition of contextual regret (Def 5) for each player when the game rewards follow the TW design rule (6). At this point, we can use property *(iii)* of Lemma 3 to obtain:

$$\sum_{t=1}^{T}\gamma(\boldsymbol{\pi}_\star(z_t)+\mathbf{x}_t,z_t)-\underbrace{\sum_{t=1}^{T}\gamma(\boldsymbol{\pi}_\star(z_t),z_t)}_{\mathrm{OPT}_c}\geq\big(1-c_{\mathrm{wc}}(\{z_t\}_{t=1}^T)\big)\big[\sum_{t=1}^{T}\gamma(\mathbf{x}_t,z_t)-\sum_{t=1}^{T}\gamma(\mathbf{0},z_t)\big].$$

The proof is completed applying the bound above to (40) and rearranging the terms. □

2) Under ES design, the same proof steps as in Proof of Thm 2 (Appendix A.4) lead to,

$$\begin{aligned}
\sum_{t=1}^{T}\gamma(\mathbf{x}_t,z_t) &\geq \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r:x_t^i[r]>0}\frac{1}{|\mathbf{x}_t|_r}\cdot\gamma^r(\mathbf{x}_t,z_t)\\
\text{(Lemma 1)}\quad &\geq \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r:x_t^i[r]>0}\frac{1}{|\mathbf{x}_t|_r}\cdot\mathrm{ucb}_t^r(\mathbf{x}_t,z_t)-2\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r:x_t^i[r]>0}\frac{1}{|\mathbf{x}_t|_r}\cdot\beta_t^r\sigma_t^r(\mathbf{x}_t,z_t)\\
\text{(Def. of Contextual Regret)}\quad &\geq \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r:\pi_\star^i(z_t)[r]>0}\frac{1}{|(\pi_\star^i(z_t),x_t^{-i})|_r}\cdot\mathrm{ucb}_t^r(\pi_\star^i(z_t),x_t^{-i},z_t)-\sum_{t=1}^{T}2\sigma_t-\sum_{i=1}^{N}R_c^i(T) \qquad (41)\\
\text{(Lemma 1)}\quad &\geq \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r:\pi_\star^i(z_t)[r]>0}\frac{1}{|(\pi_\star^i(z_t),x_t^{-i})|_r}\cdot\gamma^r(\pi_\star^i(z_t),x_t^{-i},z_t)-\sum_{t=1}^{T}2\sigma_t-\sum_{i=1}^{N}R_c^i(T)\\
\text{(Lemma 4)}\quad &\geq \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{r:\pi_\star^i(z_t)[r]>0}\gamma^r(\pi_\star^i(z_t),x_t^{-i},z_t)-\gamma(0,x_t^{-i},z_t)-\epsilon^r(z_t)-\sum_{t=1}^{T}2\sigma_t-\sum_{i=1}^{N}R_c^i(T)\\
&\geq \sum_{t=1}^{T}\sum_{i=1}^{N}\gamma(\pi_\star^i(z_t),x_t^{-i},z_t)-\gamma(0,x_t^{-i},z_t)-\sum_{t=1}^{T}2\sigma_t-\sum_{i=1}^{N}R_c^i(T)-N\sum_{t=1}^{T}\sum_{r=1}^{R}\epsilon^r(z_t),
\end{aligned}$$

where in (41) we have used the definition of contextual regret (Def 5) under ES design (8). Then, the proof is concluded by lower bounding the first summation in the bound above as it was done for case *1)* after equation (39).

□

C. Supplementary material for Section 6

Monotonicity and DR-submodularity of the considered objective. We formally show that for any context z_t and any region r, the number of daily trips (according to our simulator, see Section 6) starting from r, $\gamma^r(\cdot, z_t)$, is a monotone DR-submodular function. Consider two possible allocations $\mathbf{x}_1, \mathbf{x}_2$ with $\mathbf{x}_1 \leq \mathbf{x}_2$, i.e., under $\mathbf{x}_2$ there exists a region where at least one more bike is dropped compared to $\mathbf{x}_1$. Then, monotonicity of $\gamma^r(\cdot, z_t)$ simply follows from the fact that all trips resulting from allocation $\mathbf{x}_1$ would also be successfully completed under allocation $\mathbf{x}_2$, because there are at least the same number of available bikes per region at any point during the day. DR-submoduarity can be proved as follows. Consider allocation $\mathbf{x}_1$ and imagine an extra bike is dropped into the system at region $\bar{r}$. The increase in the number of daily trips, i.e., $\gamma^r(\mathbf{x}_1 + \mathbf{e}_{\bar{r}}, z_t) - \gamma^r(\mathbf{x}_1, z_t)$ coincides with the number of trips that utilize such extra bike, assuming that such bike is used only when no other bike is available in the same region. This number is greater than $\gamma^r(\mathbf{x}_2 + \mathbf{e}_{\bar{r}}, z_t) - \gamma^r(\mathbf{x}_2, z_t)$, since under $\mathbf{x}_2$ at least the same number of bikes is available in each region at any point in time compared to $\mathbf{x}_1$.

All the computations were carried on a 16Gb machine at 3.1 GHz. Computation times per iteration of D-SUBUCB under ES design are plotted in Figure 5 below (they are governed by the RKHS regression complexity which scales as $\mathcal{O}(t^3)$, and are similar under TW design). The large variance in CPU time across consecutive iterations is due to using two distinct models for weekdays and weekends, respectively.

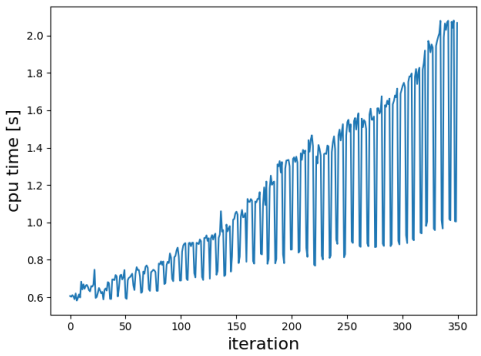

Figure 5. CPU times of D-SUBUCB under ES design