1 Proofs

1.1 DR-submodularity

In order to use the Frank-Wolfe algorithm, we must verify that the objective function is *diminishing returns* (DR) submodular. While every submodular set function is also diminishing returns, this is an additional property which must be separately verified for continuous function. Since we have already showed that G is submodular, showing that it is DR-submodular amounts to verifying that $\frac{\partial^2 G}{\partial \nu_i^2} \leq 0$ always holds (Bian et al. 2017). We again focus on a single term in the posynomial expression, $\prod_{i=1}^n (1-\nu_i)^{p_{ij}}$. If $p_{ij} \leq 1$, we are done. Assume that $p_{ij} > 1$. Taking derivatives yields

$$\frac{\partial^2}{\partial \nu_i^2}\left[\prod_{i=1}^n (1-\nu_i)^{p_{ij}}\right] = p_{ij}(p_{ij}-1)(1-\nu_i)^{p_{ij}-2}\prod_{k\neq i}(1-\nu_k)^{p_{kj}}.$$

Each term in this expression is nonnegative. We conclude that $\frac{\partial^2 F}{\partial \nu_i^2} \geq 0$ and hence $\frac{\partial^2 G}{\partial \nu_i^2} \leq 0$.

1.2 Deterministic case

We now prove the approximation guarantee for the Frank-Wolfe algorithm applied to Problem 1 (the deterministic problem).

Proof of Theorem 1. The major step is to show that ∇G is Lipschitz. Since the ℓ_∞ and ℓ_i norms are dual, it suffices to bound $||\nabla^2 H||_\infty$ to show that ∇G is Lipschitz with respect to the ℓ_1 norm. A naive bound would depend on the sizes of the coefficients in the objective, e.g., the population size and so on. We can get around this by considering (purely for the sake of analysis) the rescaled function $\frac{G(\nu)}{|G(\nu^*)|}$. Since $\frac{1}{|G(\nu^*)|}$ is a nonnegative constant with respect to ν, maximizing the two functions is equivalent. Note that if we have a guarantee of the form $\frac{G(\nu)}{|G(\nu^*)|} \geq \left(1-\frac{1}{e}\right)\frac{G(\nu^*)}{|G(\nu^*)|} - \epsilon$ then this implies that $G(\nu) \geq \left(1-\frac{1}{e}-\epsilon\right)G(\nu^*)$. That is, we get a multiplicative guarantee with respect to the unscaled function. We now consider a single element of $\nabla^2 G(\nu)$ using the posynomial representation of G. We emphasize that this representation is used purely for analysis; it does not need to be known or computed by the algorithm.

$$\begin{aligned}\left|\frac{\partial^2 G(\nu)}{\partial \nu_i^2}\right| &= \sum_j a_j p_{ij}(p_{ij}-1)(1-\nu_i)^{-2}\prod_k (1-\nu_k)^{p_{kj}} \\ &\leq \left(\frac{T}{1-U_i}\right)^2 \sum_j a_j \prod_k (1-\nu_k)^{p_{kj}} \\ &\leq \left(\frac{T}{1-U_i}\right)^2 |G(\nu^*)|\end{aligned}$$

From which we conclude that $\left|\left|\nabla^2 \frac{G(\nu)}{|G(\nu^*)|}\right|\right|_\infty \leq \left(\frac{T}{1-U_{max}}\right)^2$, where $U_{max} = \max_i U_i$. In order to apply the result of Bian et al. we actually need a bound on the Lipschitz constant of the single-dimensional auxiliary function $g_{\nu,y}(\delta) = G(\nu+\delta y)$ for any feasible ν, feasible y, and $\delta \geq 0$. Note that given any δ_1, δ_2, we have $||(\nu+\delta_1 y)-(\nu+\delta_2 y)||_1 = |\delta_1-\delta_2|\cdot||y||_1$. From this and the Lipschitz bound on G, we obtain that

$$\begin{aligned}|g_{\nu,y}(\delta_1) - g_{\nu,y}(\delta_2)| &\leq |\delta_1 - \delta_2| \left(\frac{T}{1-U_{max}}\right)^2 ||y||_1 \\ &\leq |\delta_1 - \delta_2| \left(\frac{T}{1-U_{max}}\right)^2 K\end{aligned}$$

So we have that for any ν and y, $g_{\nu,y}$ is $K\left(\frac{T}{1-U_{max}}\right)^2$-Lipschitz. Corollary 1 of Bian et al. now implies that by taking $\frac{K}{2\epsilon}\left(\frac{T}{1-U_{max}}\right)^2$ iterations in the Frank-Wolfe algorithm the guarantee in the theorem follows. □

1.3 Stochastic case

We now prove our approximation guarantee for the stochastic setting. We start out by establishing a useful smoothness property for G, which states that G is close to its linearization over small step sizes. To do so, we include for completeness the following technical relation between an ℓ_∞ bound on a function's norm and Lipschitz smoothness in the ℓ_1 norm.

Lemma 1. *Consider a differentiable function $f: R^n \to R^n$. If $||\nabla f(x)||_\infty \leq L \; \forall x \in \mathcal{P}$, $||f(y) - f(x)||_1 \leq L||y-x||_1 \; \forall x, y \in \mathcal{P}$.*

Proof. Fix any $x, y \in \mathcal{P}$. Define the auxiliary function $g(\delta) = f(x + \delta(y-x))$ We have $f(y) - f(x) = g(1) - g(0)$ and hence

$$\begin{aligned}||f(y) - f(x)||_1 &= ||g(1) - g(0)||_1 \\ &= \left|\left|\int_0^1 \frac{dg(\delta)}{d\delta}\, d\delta\right|\right|_1 \\ &\leq \int_0^1 \left|\left|\frac{dg(\delta)}{d\delta}\right|\right|_1 d\delta \\ &= \int_0^1 \left|\left|\nabla f(x + \delta(y-x))^\top (y-x)\right|\right|_1 d\delta \\ &\leq \int_0^1 ||\nabla f(x + \delta(y-x))^\top||_\infty ||y-x||_1 \text{ (Hölder's inequality)} \\ &\leq L||y-x||_1\end{aligned}$$

□

Lemma 2. *Suppose that G has an L-Lipschitz gradient in the ℓ_1 norm. Let $d = \max_{x \in \mathcal{P}} ||x||_\infty ||x||_1$. For any x, y, $G(x + \gamma y) - G(x) \geq \gamma \nabla G(x)^T y - \frac{Ld\gamma^2}{2}$.*

Proof. For any $x, y \in \mathcal{P}$, we consider the one dimensional auxiliary function $g_{x,y}(\delta) = G(x + \delta y)$. We can show that g has a gradient which is Ld-Lipschitz:

$$
\begin{aligned}
\frac{dg(\delta_1)}{d\delta} - \frac{dg(\delta_2)}{d\delta} &= \nabla G(x+\delta_1 y)^\top y - \nabla G(x+\delta_2 y)^\top y \\
&= \left(\nabla G(x+\delta_1 y) - \nabla G(x+\delta_2 y)\right)^\top y \\
&\leq ||\nabla G(x+\delta_1 y) - \nabla G(x+\delta_2 y)||_1 ||y||_\infty \text{ (by Hölder's inequality)} \\
&\leq L||(x+\delta_1 y) - (x+\delta_2) y||_1 ||y||_\infty \\
&\leq L|\delta_1 - \delta_2| \cdot ||y||_1 ||y||_\infty \\
&= Ld|\delta_1 - \delta_2|.
\end{aligned}
$$

Now, we use smoothness of g to establish that G is close to its linearization over short distances:

$$
\begin{aligned}
G(x+\gamma y) - G(x) - \nabla G(x)^T y &= g(\gamma) - g(0) - \frac{dg(0)}{d\delta} \cdot 1 \\
&= \int_{\delta=0}^{\gamma} \left[\frac{dg(\delta)}{d\delta} - \frac{dg(0)}{d\delta}\right] d\delta \\
&\leq \int_{\delta=0}^{\gamma} Ld\delta \, d\delta \\
&= \frac{Ld\gamma^2}{2}.
\end{aligned}
$$

which proves the lemma. □

We also use the following lemma, the proof of which can be found in Bian et al. (2017):

Lemma 3. *For any DR-submodular function G and its optimizer ν^*, $G(\nu^*+\nu) - G(\nu) \leq \nabla G(\nu)^\top \nu^*$.*

Using these lemmas, we prove the following general result for any smooth DR-submodular function:

Theorem 1. *Let G be a DR-submodular function which is C-Lipschitz in the ℓ_1 norm, with L-Lipschitz gradient (also in ℓ_1 norm) and $G(0) = 0$. Let $d = \max_{x\in\mathcal{P}} ||x||_\infty ||x||_1$ and $b = \max_{x\in\mathcal{P}} ||x||_1$. Then, running DOMO for $R = \frac{Ld}{\epsilon}$ iterations with $m = \left(\frac{4bC}{\epsilon}\right)^2$ samples per iteration yields a feasible $\nu \in \mathcal{P}$ which satisfies $\mathbb{E}[G(\nu)] \geq (1-\frac{1}{e})G(\nu^*) - \epsilon$.*

Proof of Theorem 2. We consider a DR-submodular function with L-Lipschitz gradient. Further, we assume that G itself is C-Lipschitz. We analyze the gain made in a single step as follows:

$$
\begin{aligned}
G(\nu^k) - G(\nu^{k-1}) &\geq \frac{1}{R}\nabla G(\nu^{k-1})^\top y^k - \frac{Ld}{2R^2} \text{ (Lemma 2)}\\
&= \frac{1}{R}\widehat{\nabla}_k^\top y^k - \frac{1}{R}\left(\widehat{\nabla}_k - \nabla G(\nu^{k-1})\right)^\top y^k - \frac{Ld}{2R^2}\\
&\geq \frac{1}{R}\widehat{\nabla}_k^\top y^k - \frac{b}{R}||\widehat{\nabla}_k - \nabla G(\nu^{k-1})||_\infty - \frac{Ld}{2R^2} \text{ (by Hölder's inequality and } ||y||_1 \leq b)\\
&\geq \frac{1}{R}\widehat{\nabla}_k^\top \nu^* - \frac{b}{R}||\widehat{\nabla}_k - \nabla G(\nu^{k-1})||_\infty - \frac{Ld}{2R^2} \text{ (by definition of } y^k)\\
&= \frac{1}{R}\nabla G(\nu^{k-1})^\top \nu^* - \frac{1}{R}\left(G(\nu^{k-1}) - \widehat{\nabla}_k\right)^\top \nu^* - \frac{b}{R}||\widehat{\nabla}_k - \nabla G(\nu^{k-1})||_\infty - \frac{Ld}{2R^2}\\
&\geq \frac{1}{R}\nabla G(\nu^{k-1})^\top \nu^* - \frac{2b}{R}||\widehat{\nabla}_k - \nabla G(\nu^{k-1})||_\infty - \frac{Ld}{2R^2}\\
&\geq \frac{1}{R}\left(G(\nu^* + \nu^{k-1}) - G(\nu^{k-1})\right) - \frac{2b}{R}||\widehat{\nabla}_k - \nabla G(\nu^{k-1})||_\infty - \frac{Ld}{2R^2} \text{ (Lemma 3)}\\
&\geq \frac{1}{R}\left(G(\nu^*) - G(\nu^{k-1})\right) - \frac{2b}{R}||\widehat{\nabla}_k - \nabla G(\nu^{k-1})||_\infty - \frac{Ld}{2R^2} \text{ (monotonicity)}
\end{aligned}
$$

By assumption that G is C-Lipschitz, $||\nabla G||_\infty \leq C$. Via Jensen's inequality, we have

$$
\begin{aligned}
\mathbb{E}\left[||\widehat{\nabla}_k - \nabla G(\nu^{k-1})||_\infty\right] &\leq \sqrt{\mathbb{E}\left[||\widehat{\nabla}_k - \nabla G(\nu^{k-1})||_\infty^2\right]}\\
&\leq \frac{C}{\sqrt{m}}
\end{aligned}
$$

where the last step uses that averaging over m independent samples reduces the variance of $\widehat{\nabla}_k$ by a factor of m. Hence we have

$$
\mathbb{E}\left[G(\nu^*) - G(\nu^k)\right] \leq \left(1 - \frac{1}{R}\right)\mathbb{E}\left[G(\nu^*) - G(\nu^{k-1})\right] - \frac{2bC}{R\sqrt{m}} - \frac{Ld}{2R^2}.
$$

and so

$$
\begin{aligned}
\mathbb{E}\left[G(\nu^*) - G(\nu^R)\right] &\leq \left(1 - \frac{1}{R}\right)^R \mathbb{E}\left[G(\nu^*) - G(\nu^0)\right] - \sum_{k=0}^{R-1}\frac{2bC}{R\sqrt{m}} - \sum_{k=0}^{R-1}\frac{Ld}{2R^2}\\
&\leq \left(1 - \frac{1}{R}\right)^R \mathbb{E}\left[G(\nu^*) - G(\nu^0)\right] - \frac{2bC}{\sqrt{m}} - \frac{Ld}{2R}
\end{aligned}
$$

and hence

$$
\mathbb{E}\left[G(\nu^*) - G(\nu^R)\right] \leq \frac{1}{e}\left[G(\nu^*) - G(0)\right] - \frac{2bC}{\sqrt{m}} - \frac{Ld}{2R}.
$$

Choosing $m = \left(\frac{4bC}{\epsilon}\right)^2$ and $R = \frac{Ld}{\epsilon}$ completes the proof. □

Now, the result for our problem follows by noting the appropriate values for the Lipschitz constant and size of the feasible set. Note that the stochastic objective $H := \mathbb{E}[G(\cdot, \xi)]$ inherits the smoothness properties of the deterministic objective since it is a convex combination of such functions. Using the same reasoning as the deterministic case, the rescaled objective $\frac{H}{H(\nu^*)}$ satisfies $\left|\left|\nabla \frac{H}{H(\nu^*)}\right|\right|_\infty \leq \frac{T}{(1-U_{max})}$ and $\left|\left|\nabla^2 \frac{H}{H(\nu^*)}\right|\right|_\infty \leq \left(\frac{T}{1-U_{max}}\right)^2$. Via Lemma 1, this yields immediate bounds for the Lipschitz constants C and L. Moreover, $d = b = K$. Plugging these values into Theorem 1 yields the result.

2 Experiments

We now provide additional detail about the data sources used in the experiments.

2.1 TB

- Annual death probabilities μ: Calculated from WHO life tables (available at `http://apps.who.int/gho/data/?theme=main&vid=61780`)
- Total population N: reported in United Nations Statistics Division Demographic Statistics (available at `http://data.un.org/Data.aspx?d=POP&f=tableCode%3A22`)
- Clearance rate ν: Taken from RNTCP Annual Performance Reports (available at `http://tbcindia.gov.in/index4.php?lang=1&level=0&linkid=380&lid=2746`).
- Most individuals with untreated TB will die within three years[1]. We set the annual probability of death for TB patients such that 90% will die after 3 years of infection.

2.2 Gonorrhea

- Estimates of prevalence I: U.S. Centers for Disease Control. `https://www.cdc.gov/std/stats15/tables/21.htm` Table 21. Gonorrhea Reported Cases and Rates of Reported Cases by Age Group and Sex, United States, 2011-2015.
- Annual probability of death μ: Calculated from WHO life tables (available at `http://apps.who.int/gho/data/?theme=main&vid=61780`)
- Total population N: Taken from US Census Data (Annual estimates of the resident population by single year of age and sex for the United States: April 1, 2010 to July 1, 2016 (NC-EST2016-AGESEX-RES), available online at `https://www.census.gov/data/datasets/2016/demo/popest/nation-detail.html`)
- Once detected, most cases of gonorrhea can be effectively treated within two or three months, so the national clearance rate for this annual model depends on the detection rate of the disease. The clearance rate therefore varies over the samples used to calculate the transmission matrix.
- We assume no one dies from gonorrhea. Death rates are similar to the non-disease death rates.

[1] Tiemersma EW, van der Werf MJ, Borgdorff MW, Williams BG, Nagelkerke NJD. Natural History of Tuberculosis: Duration and Fatality of Untreated Pulmonary Tuberculosis in HIV Negative Patients: A Systematic Review. Pai M, ed. PLoS ONE. 2011;6(4):e17601. doi:10.1371/journal.pone.0017601.