Supplemental Material: Risk-sensitive submodular optimization

Bryan Wilder
Department of Computer Science and Center for Artificial Intelligence in Society
University of Southern California
bwilder@usc.edu

1 Proofs for continuous submodular setting

Lemma 1. *Take* $s = \frac{2nM^2}{\epsilon^2} \log \frac{1}{\delta} \log \frac{L_1}{\epsilon}$ *samples and let* $\widehat{CVaR}_\alpha$ *be the empirical CVaR on the samples. Then,* $|CVaR_\alpha(\boldsymbol{x}) - \widehat{CVaR}_\alpha(\boldsymbol{x})| \leq \epsilon$ *holds for all* $\boldsymbol{x} \in \mathcal{P}$ *with probability at least* $1 - \delta$.

Proof. We can establish the result for fixed $\boldsymbol{x}$ using the proof of Ohsaka and Yoshida. We have via taking $c = L$ in their Lemma 4.4 that for any fixed $\boldsymbol{x}$, $|\text{CVaR}_\alpha(\boldsymbol{x}) - \widehat{\text{CVaR}_\alpha}(\boldsymbol{x})| \leq \epsilon$ with probability at least $1-\delta$ by taking $s = \Theta\left(\frac{M^2}{\epsilon^2}\log\frac{1}{\delta}\right)$ samples. Note that we cannot directly take union bound because the set of $x \in \mathcal{P}$ is not finite. Instead, we take a uniform grid of $\left(\frac{L_1 d}{\epsilon}\right)^n$ points containing $\mathcal{P}$. Via union bound, concentration holds for all points in the grid using $s = \Theta\left(\frac{M^2}{\epsilon^2}\log\left(\left(\frac{L_1 d}{\epsilon}\right)^n \frac{1}{\delta}\right)\right) = \Theta\left(\frac{M^2 n}{\epsilon^2}\log\frac{L_1 d}{\epsilon\delta}\right)$. Now we argue that every point in $\mathcal{P}$ is close in CVaR value to a point in the grid. The grid has enough points to guarantee that for any $\boldsymbol{x}_1 \in \mathcal{P}$, there is a point $\boldsymbol{x}_2 \succeq \boldsymbol{x}_1$ within ℓ_2 distance $\frac{\epsilon}{L_2}$ of $\boldsymbol{x}_1$. Note that $\text{CVaR}_\alpha(\boldsymbol{x}_2) \geq \text{CVaR}_\alpha(\boldsymbol{x}_1)$ by monotonicity of F combined with monotonicity of CVaR. Additionally by monotonicity of CVaR, $\text{CVaR}_\alpha(\boldsymbol{x}_2) - \text{CVaR}_\alpha(\boldsymbol{x}_1)$ is maximized when for all y, $F(\boldsymbol{x}_2, y) = F(\boldsymbol{x}_1, y) + L||\boldsymbol{x}_1 - \boldsymbol{x}_2||_2$. In this case, $\{y|F(\boldsymbol{x}_1) \leq \text{VaR}_\alpha(\boldsymbol{x}_1)\} = \{y|F(\boldsymbol{x}_2) \leq \text{VaR}_\alpha(\boldsymbol{x}_2)\}$. Let $\mathcal{Z}$ denote this set of scenarios. We have $\text{CVaR}_\alpha(\boldsymbol{x}_1) = \mathbb{E}\left[F(\boldsymbol{x}_1, y)|y \in \mathcal{Z}\right]$ and $\text{CVaR}_\alpha(\boldsymbol{x}_2) = \mathbb{E}\left[F(\boldsymbol{x}_2, y)|y \in \mathcal{Z}\right]$. But since we take the expectation over a fixed set of scenarios and each $F(\cdot, y)$ is L_1-Lipschitz, the expectation must be L_1-Lipschitz as well. Hence, $\text{CVaR}_\alpha(\boldsymbol{x}_2) - \text{CVaR}_\alpha(\boldsymbol{x}_1) \leq L_1||\boldsymbol{x}_1 - \boldsymbol{x}_2||_2 \leq \epsilon$. $\square$

Lemma 2. *Define* $g(\tau) = \sum_{y \in \mathcal{Y}} I_y(\tau)$. *(a)* τ *maximizes* $\frac{1}{u}\int_{z=0}^{u} H(\boldsymbol{x}, \tau) dz$ *if* $g(\tau) = \alpha s$. *(b)* g *is piecewise linear and monotone decreasing.*

Proof. We start with the claim in (a). We have

$$\frac{1}{u}\int_{z=0}^{u} H(\boldsymbol{x}, \tau) dz = \frac{1}{u}\int_{z=0}^{u} \tau - \frac{1}{\alpha s}\sum_{F(\boldsymbol{x}, y) \leq \tau + z} \tau - F(\boldsymbol{x}, y) dz.$$

Note that the function inside the integral is known to be concave in τ (Rockafellar and Urseyev 2000), which yields concavity of the entire function. Thus, to find a maximum it suffices to find a point where the derivative with respect to τ is 0. To this end, note that the set of z such that $F(\boldsymbol{x}, y) = \tau + z$ for some z has measure 0 and hence do not impact the value of the integral. For the remaining values of z, we have

$$\frac{d}{d\tau} = 1 - \frac{|\{y : F(\boldsymbol{x}, y) \leq \tau + z\}|}{\alpha s}$$

since the set in the numerator is constant over some interval around τ. This yields

$$\begin{aligned}\frac{d}{dt}\frac{1}{u}\int_{z=0}^{u} H(\boldsymbol{x},\tau)dz &= \frac{1}{u}\int_{z=0}^{u} 1-\frac{|\{y : F(\boldsymbol{x},y)\leq \tau+z\}|}{\alpha s}dz \\ &= \frac{1}{u}\left[1-\frac{1}{\alpha s}\sum_{y\in\mathcal{Y}}\int_{z=0}^{1}\mathbb{1}\left[F(\boldsymbol{x},y)\leq\tau\right]dz\right] \\ &= \frac{1}{u}\left[1-\frac{1}{\alpha s}\sum_{y\in\mathcal{Y}} I_y(\tau)\right].\end{aligned}$$

By inspection, the derivative is 0 when $\sum_{y\in\mathcal{Y}} I_y(\tau)=\alpha s$, which proves part (a) of the lemma.
For part (b), we simply note that each $I_y(\tau)$ is monotone decreasing and piecewise linear in τ, and g is the sum of such functions.

$\square$

Lemma 3. *For any* $\boldsymbol{x}$ *and* τ, $\left|\tilde{H}(\boldsymbol{x},\tau)-H(\boldsymbol{x},\tau)\right| \leq \frac{u\left(1+\frac{1}{\alpha}\right)}{2}$

Proof. We start out by showing that H is $\left(1+\frac{1}{\alpha}\right)$ −Lipschitz in τ. Consider any $\boldsymbol{x}$, τ, and τ' and without loss of generality let $\tau' > \tau$.

$$H(\boldsymbol{x},\tau)-H(\boldsymbol{x},\tau') = \left[\tau-\tau'\right]-\frac{1}{\alpha|\mathcal{Y}|}\sum_{y}\max(\tau-F(\boldsymbol{x},y),0)-\max(\tau'-F(\boldsymbol{x},y),0).$$

We consider three cases for the term inside the summation. First, $F(\boldsymbol{x},y)<\tau$. Here, $\max(\tau-F(\boldsymbol{x},y),0)-\max(\tau'-F(\boldsymbol{x},y),0)=\tau-\tau'$. Second, $\tau\leq F(\boldsymbol{x},y)<\tau'$. Here,

$$\max(\tau-F(\boldsymbol{x},y),0)-\max(\tau'-F(\boldsymbol{x},y),0)=F(\boldsymbol{x},y)-\tau'$$

and hence

$$\left|\max(\tau-F(\boldsymbol{x},y),0)-\max(\tau'-F(\boldsymbol{x},y),0)\right|\leq|\tau-\tau'|.$$

Third, $F(\boldsymbol{x},y)\geq\tau'$. Here, the term in the summation is zero.
Via the triangle inequality, we conclude that

$$\begin{aligned}\left|H(\boldsymbol{x},\tau)-H(\boldsymbol{x},\tau')\right| &\leq \left|\tau-\tau'\right|+\frac{1}{\alpha|\mathcal{Y}|}\sum_{y}|\tau-\tau'| \\ &\leq \left(1+\frac{1}{\alpha}\right)|\tau-\tau'|.\end{aligned}$$

Now, since $z\in[0,u]$ holds with probability 1, we can apply the above reasoning to conclude that

$$
\begin{aligned}
\left|\tilde{H}(\boldsymbol{x},\tau)-H(\boldsymbol{x},\tau)\right| &= \frac{1}{u}\left|\int_{z=0}^{u} H(\boldsymbol{x},\tau+z)-H(\boldsymbol{x},\tau)dz\right| \\
&\leq \frac{1}{u}\int_{z=0}^{u} \left|H(\boldsymbol{x},\tau+z)-H(\boldsymbol{x},\tau)\right| dz \\
&\leq \frac{1}{u}\int_{z=0}^{u} \left(1+\frac{1}{\alpha}\right) z dz \\
&= \frac{u\left(1+\frac{1}{\alpha}\right)}{2}
\end{aligned}
$$

□

Lemma 4. *At each iteration $k = 1...K$, $\tilde{H}(\tilde{\boldsymbol{x}}^*, \tau(\tilde{\boldsymbol{x}}^*)) - \tilde{H}(\boldsymbol{x}^k, \tau(\boldsymbol{x}^k)) \leq \max_{\boldsymbol{v}\in\mathcal{P}}\langle \nabla_{\boldsymbol{x}} \tilde{H}(\boldsymbol{x}^k, \tau(\boldsymbol{x}^k)), \boldsymbol{v}\rangle$.*

Proof. We will show that $\max_\tau \tilde{H}(\cdot, \tau)$ is an up-concave function. Fix any two points $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{P}$. We start out by defining the function $h : [0, 1] \to R$ as $h(\xi) = H(\boldsymbol{x}_1 + \xi\boldsymbol{x}_2)$. We will show that $h(\xi, \tau)$ is jointly concave in (ξ, τ). To show joint concavity in (ξ, τ), we write

$$
h(\xi,\tau) = \tau - \frac{1}{\alpha|\mathcal{Y}|}\sum_{y\in\mathcal{Y}} [\tau - F(\boldsymbol{x}_1 + \xi\boldsymbol{x}_2, y)]^+
$$

The first term is linear in (ξ, τ), and so is concave. We will show that the expectation is jointly convex in (ξ, τ), from which concavity of h follows. To show this, is suffices to show that the term inside the expectation is convex for any fixed y. Note that this term is the composition of the function $(\xi, \tau) \mapsto \tau - F(\boldsymbol{x}_1 + \xi\boldsymbol{x}_2, y)$ with the function $t \mapsto [t]^+$. Since the latter is a nondecreasing convex function, the composition is convex whenever the inner function is convex. For the inner function, τ is convex because it is linear in ξ, and $-F(\boldsymbol{x}_1 + \xi\boldsymbol{x}_2, y)$ is convex because $F(\cdot, y)$ is up-concave. Thus, the claim follows.

Now define $\tilde{h}(\xi, \tau) = \tilde{H}(\boldsymbol{x}_1 + \xi\boldsymbol{x}_2)$. $\tilde{h}$ is jointly concave in (ξ, τ) because it is a nonnegative linear combination of concave functions. This shows that $\max_\tau \tilde{h}(\xi, \tau)$ is concave in ξ because maximizing a jointly concave function with respect to one of its parameters yields a concave function in the remaining parameters. We conclude that $\max_\tau \tilde{H}(\cdot, \tau)$ is an up-concave function.

Fix the points $\boldsymbol{x}^k$ and $\tilde{\boldsymbol{x}}^*$ as $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in the above definition of $\tilde{h}$. Now the conclusion follows by arguing (as in Bian et al. (2017)),

$$
\begin{aligned}
\tilde{H}(\tilde{\boldsymbol{x}}^*, \tau(\tilde{\boldsymbol{x}}^*)) - \tilde{H}(\boldsymbol{x}^k, \tau(\boldsymbol{x}^k)) &\leq \tilde{H}(\tilde{\boldsymbol{x}}^* + \boldsymbol{x}^k, \tau(\tilde{\boldsymbol{x}}^* + \boldsymbol{x}^k)) - \tilde{H}(\boldsymbol{x}^k, \tau(\boldsymbol{x}^k)) \\
&= \max_\tau \tilde{h}(1,\tau) - \max_\tau \tilde{h}(0,\tau) \\
&\leq \frac{d \max_\tau \tilde{h}(0,\tau)}{d\xi} \cdot 1 \text{ (since } \max_\tau \tilde{h}(\cdot,\tau)\text{) is concave} \\
&= \langle \nabla_{\boldsymbol{x}} \tilde{H}(\boldsymbol{x}^k, \tau(\boldsymbol{x}^k)), \tilde{\boldsymbol{x}}^*\rangle \\
&\leq \max_{\boldsymbol{v}\in\mathcal{P}} \langle \nabla_{\boldsymbol{x}} \tilde{H}(\boldsymbol{x}^k, \tau(\boldsymbol{x}^k)), \boldsymbol{v}\rangle
\end{aligned}
$$

□

In the lemmas that follow, we use an alternate interpretation of $\tilde{H}$. Namely, we can view the smoothing process as drawing a random variable z from a uniform distribution over the interval $[0, u]$. Then, $\tilde{H}(\boldsymbol{x}, \tau) = \mathbb{E}_z [H(\boldsymbol{x}, \tau + z)]$. This is completely equivalent to the definition given in the text, but simplifies notation and concepts at a few places in the proofs below.

Lemma 5. *If $\boldsymbol{x}_2 \succeq \boldsymbol{x}_1$, $\nabla_{\boldsymbol{x}} \tilde{H}(\boldsymbol{x}_2, \tau(\boldsymbol{x}_2)) \succeq \nabla_{\boldsymbol{x}} \tilde{H}(\boldsymbol{x}_2, \tau(\boldsymbol{x}_1))$.*

Proof. Recall that $\nabla_{\boldsymbol{x}} \tilde{H}(\boldsymbol{x}_2, \tau) = \mathbb{E}_z [\nabla_{\boldsymbol{x}} H(\boldsymbol{x}_2, \tau + z)]$. We couple the random variables $\nabla_{\boldsymbol{x}} H(\boldsymbol{x}_2, \tau(\boldsymbol{x}_1) + z)$ and $\nabla_{\boldsymbol{x}} H(\boldsymbol{x}_2, \tau(\boldsymbol{x}_2) + z)$ by fixing z to any value where both derivatives exist (which excludes only a measure 0 set).

Note that since F is monotone in $\boldsymbol{x}$, $\nabla_{\boldsymbol{x}} F(\boldsymbol{x}, y) \succeq 0$ holds for all $\boldsymbol{x} \in \mathcal{P}$ and $y \in \mathcal{Y}$. Moreover, we can write

$$\nabla_{\boldsymbol{x}} H(\boldsymbol{x}, \tau) = \frac{1}{\alpha|\mathcal{Y}|} \sum_{y \in \mathcal{Y}: F(\boldsymbol{x}, y) \leq \tau} \nabla_{\boldsymbol{x}} F(\boldsymbol{x}, y).$$

It is easy to see that the function $\tau(\boldsymbol{x})$ is monotone nondecreasing and hence $\tau(\boldsymbol{x}_2) + z \geq \tau(\boldsymbol{x}_1) + z$. Thus, for all $y \in \mathcal{Y}$, $F(\boldsymbol{x}_2) \leq \tau(\boldsymbol{x}_2) + z$ only if $F(\boldsymbol{x}_2) \leq \tau(\boldsymbol{x}_1) + z$. Since each term in the above summation is nonnegative, $\nabla_{\boldsymbol{x}} H(\boldsymbol{x}_2, \tau(\boldsymbol{x}_1) + z) \succeq \nabla_{\boldsymbol{x}} H(\boldsymbol{x}_2, \tau(\boldsymbol{x}_2) + z)$. The lemma now follows by taking the expectation with respect to z. $\square$

Lemma 6. *If $\forall y \in \mathcal{Y}$, $F(\cdot, y)$ is L_1-Lipschitz and $\nabla_{\boldsymbol{x}} F(\cdot, y)$ is L_2 Lipschitz with $||\nabla_{\boldsymbol{x}} F||_2 \leq G$, then $\nabla_{\boldsymbol{x}} \tilde{H}$ is $\frac{1}{\alpha}\left(L_2 + \frac{L_1 G}{u}\right)$ $-$Lipschitz.*

Proof. For any z, let $\mathcal{Y}_1(z) = \{y : F(\boldsymbol{x}_1, y) \leq F(\boldsymbol{x}_2, y) \leq \tau + z\}$. Let $\mathcal{Y}_2(z) = \{y : F(\boldsymbol{x}_1, y) \leq \tau + z < F(\boldsymbol{x}_2, y)\}$. We have

$$\begin{aligned}
||\nabla_{\boldsymbol{x}} \tilde{H}(\boldsymbol{x}_1, \tau) - \nabla_{\boldsymbol{x}} \tilde{H}(\boldsymbol{x}_2, \tau)|| &= \left|\left|\mathbb{E}_z [\nabla_{\boldsymbol{x}} H(\boldsymbol{x}_1, \tau + z)] - \mathbb{E}_z [\nabla_{\boldsymbol{x}} H(\boldsymbol{x}_2, \tau + z)]\right|\right| \\
&= \frac{1}{\alpha} \left|\left| \mathbb{E}_z \left[\frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}_1(z)} \nabla_{\boldsymbol{x}} F(\boldsymbol{x}_1, y) - \nabla_{\boldsymbol{x}} F(\boldsymbol{x}_2, y) - \frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}_2(z)} \nabla_{\boldsymbol{x}} F(\boldsymbol{x}_1, y) \right] \right|\right| \\
&\leq \frac{1}{\alpha} \mathbb{E}_z \left[\frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}_1(z)} ||\nabla_{\boldsymbol{x}} F(\boldsymbol{x}_1, y) - \nabla_{\boldsymbol{x}} F(\boldsymbol{x}_2, y)|| + \frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}_2(z)} ||\nabla_{\boldsymbol{x}} F(\boldsymbol{x}_1, y)|| \right] \\
&\leq \frac{1}{\alpha} \mathbb{E}_z \left[\frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}_1(z)} L_2 \, ||\boldsymbol{x}_1 - \boldsymbol{x}_2|| + \frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}_2(z)} ||\nabla_{\boldsymbol{x}} F(\boldsymbol{x}_1, y)|| \right] \\
&\leq \frac{1}{\alpha} \mathbb{E}_z \left[\frac{|\mathcal{Y}_1(z)|}{|\mathcal{Y}|} \right] L_2 ||\boldsymbol{x}_2 - \boldsymbol{x}_1|| + \frac{1}{\alpha} \mathbb{E}_z \left[\frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}_2(z)} ||\nabla_{\boldsymbol{x}} F(\boldsymbol{x}_1, y)|| \right] \\
&\leq \frac{1}{\alpha} \mathbb{E}_z \left[\frac{|\mathcal{Y}_1(z)|}{|\mathcal{Y}|} \right] L_2 ||\boldsymbol{x}_2 - \boldsymbol{x}_1|| + \frac{1}{\alpha} \mathbb{E}_z \left[\frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}_2(z)} G \right] \\
&= \frac{1}{\alpha} \mathbb{E}_z \left[\frac{|\mathcal{Y}_1(z)|}{|\mathcal{Y}|} \right] L_2 ||\boldsymbol{x}_2 - \boldsymbol{x}_1|| + \frac{1}{\alpha} \mathbb{E}_z \left[\frac{G}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} 1 \left[y \in \mathcal{Y}_2(z) \right] \right] \\
&= \frac{1}{\alpha} \mathbb{E}_z \left[\frac{|\mathcal{Y}_1(z)|}{|\mathcal{Y}|} \right] L_2 ||\boldsymbol{x}_2 - \boldsymbol{x}_1|| + \frac{G}{\alpha|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} \Pr_z \left[y \in \mathcal{Y}_2(z) \right]
\end{aligned}$$

Now, all that remains is to bound the term $\Pr_z[y \in \mathcal{Y}_2(z)]$. Note that for all $y \in \mathcal{Y}$, $F(\boldsymbol{x}_2, y) \leq F(\boldsymbol{x}_1, y) + L_1||\boldsymbol{x}_1 - \boldsymbol{x}_2||$. Since z follows a uniform distribution over an interval of size u, the probability that it falls into an interval of size $L_1||\boldsymbol{x}_1 - \boldsymbol{x}_2||$ is at most $\frac{L_1||\boldsymbol{x}_1-\boldsymbol{x}_2||}{u}$ and we conclude that

$$||\nabla_{\boldsymbol{x}}\tilde{H}(\boldsymbol{x}_1, \tau) - \nabla_{\boldsymbol{x}}\tilde{H}(\boldsymbol{x}_2, \tau)|| \leq \frac{1}{\alpha}\left(L_2 + \frac{L_1 G}{u}\right)||\boldsymbol{x}_2 - \boldsymbol{x}_1||.$$

□

2 Proofs for discrete portfolio optimization

We recall the problem setting for discrete submodular functions, which we refer to as the *discrete portfolio optimization* problem. We are given a collection of submodular set functions $f(\cdot, y)$ on a ground set X, where y is a random variable. There is a collection of feasible sets $\mathcal{I}$. For instance, $\mathcal{I}$ could be all size-k subsets. In general, we will focus on the setting where $\mathcal{I}$ forms a matroid. Analogously to the continuous setting, we assume that f is bounded: $\max_{y, S \in \mathcal{I}} f(S, y) \leq M$ for some $M > 0$. The algorithm selects a distribution q over the sets in $\mathcal{I}$. Let $\Delta(\mathcal{I})$ be the set of all such distributions (the $|\mathcal{I}|$-dimensional simplex). We aim to solve the problem

$$\max_{q \in \Delta(\mathcal{I})} \text{CVaR}_\alpha\left(\sum_{S \in \mathcal{I}} q_S f(S, y)\right)$$

In this section, we provide a block-box reduction from the above problem to the continuous submodular CVaR problem studied so far. We start by introducing a few useful concepts from submodular optimization in order to formulate our proposed algorithm. We then state the algorithm and prove its approximation guarantee. We will assume throughout that the number of scenarios y is at most the value s given in Lemma 1 since this suffices to obtain ϵ-accurate solutions to the true CVaR problem.

Multilinear extension: For a given a submodular function f, its multilinear relaxation F is a function defined over the continuous space $[0,1]^{|X|}$. For any $\boldsymbol{x} \in [0,1]^{|X|}$, let $p_{\boldsymbol{x}}$ denote the product distribution with marginals given by $\boldsymbol{x}$. We have $F(\boldsymbol{x}) = \mathbb{E}_{S \sim p_{\boldsymbol{x}}}[f(S)]$. Note that F agrees with f at each vertex of the hypercube, the points $\{0,1\}^{|X|}$ (where we interpret each binary vector as the indicator vector of a set). The value of F, as well as its gradients, can be efficiently computed via random sampling [2], with closed forms known for common special cases [4]. Here, we ignore such issues and assume that F and ∇F are available exactly (since evaluation up to arbitrary precision ϵ can be done via sampling). For any submodular function f, F is a continuous submodular function. Moreoever, F is smooth (in the sense of having Lipschitz gradients of bounded norm in terms of M; see [5]).

Swap rounding: Let $\mathcal{P}$ be the convex hull of the indicator vectors of sets in $\mathcal{I}$. Each point in $\boldsymbol{x} \in \mathcal{P}$ specifies a product distribution, and via the multilinear extension we can optimize over such distributions. However, we need to convert this product distribution back to a distribution over sets $\mathcal{I}$. Note that just sampling from $p_{\boldsymbol{x}}$ is not guaranteed to give us sets that are feasible (lie in $\mathcal{I}$). For instance, if $\mathcal{I}$ is the k-uniform matroid, sampling from a a product distribution $\boldsymbol{x} \in \mathcal{P}$ can easily produce sets with more than k elements even though $\sum_{i=1}^n x_i \leq k$. Whenever $\mathcal{I}$ is a matroid, the swap rounding algorithm of Chekuri et al. gives a means for efficiently rounding a point $\boldsymbol{x} \in \mathcal{P}$

to a single feasible set $\boldsymbol{x} \in \mathcal{P}$ in a randomized fashion. The set S satisfies $\mathbb{E}[f(S)] \geq F(\boldsymbol{x})$ for any submodular function f, and also satisfies a lower tail bound which controls the probability that $f(S)$ is significantly less than $F(\boldsymbol{x})$. Wilder [5] leverage this result to show the following:

Lemma 7. *Suppose we draw $O\left(\frac{\log \frac{N}{\delta}}{\epsilon^3}\right)$ independent samples $S_1 ... S_\ell$ via swap rounding. Then $\frac{1}{\ell}\sum_{i=1}^{\ell} f(S_i) \geq F(\boldsymbol{x}) - \epsilon$ holds for any N submodular functions and their multilinear extensions with probability at least $1-\delta$.*

Combining these two ingredients yields the following algorithm for discrete portfolio optimization:

Algorithm 1 PortfolioCVaR

Input: an algorithm $\mathcal{A}$ for continuous submodular CVaR maximization.
Set $r = O\left(\frac{M^2 \log s}{\epsilon^2}\right)$
Use $\mathcal{A}$ to solve the problem

$$\max_{\boldsymbol{x}^1 ... \boldsymbol{x}^r \in \times_{i=1}^{r} \mathcal{P}} \text{CVaR}_\alpha \left(\frac{1}{r} \sum_{i=1}^{r} \mathop{\mathbb{E}}_{S \sim p_{\boldsymbol{x}^i}} \left[f(S, y) \right] \right)$$

obtaining a solution $\boldsymbol{x}^1 ... \boldsymbol{x}^r$.
Set $\ell = O\left(\frac{\log \frac{sr}{\delta}}{\epsilon^3}\right)$
for $i = 1...r$ **do**
    Draw sets $S_1^i ... S_\ell^i$ independently as SwapRound($\boldsymbol{x}^i$).
end for
Return the uniform distribution on $\{S_j^i : i = 1...r, j = 1...\ell\}$.

This algorithm solves a continuous optimization problem using the multilinear extensions $F(\cdot, y)$. We maintain r copies of the decision variables $\boldsymbol{x}^1 ... \boldsymbol{x}^r$, a choice which is justified later (we remark that similar ideas have been used by [3, 5] in other domains). We then draw a series of samples via swap rounding for each $\boldsymbol{x}^i$ and output the uniform distribution over the combined set of samples.

Theorem 1. *Suppose that we have access to an algorithm $\mathcal{A}$ for the continuous submodular CVaR maximization problem, which returns a solution with value at least $\alpha OPT - \epsilon$ for some $\alpha > 0$ and any $\epsilon > 0$ in time poly$(M, \frac{1}{\epsilon}, n)$. Then, for any discrete portfolio optimization problem over a matroid,* PortfolioCVaR *returns a solution with value at least $\alpha OPT - \epsilon$ with probability at least $1 - \delta$ in time poly$(M, \frac{1}{\epsilon}, n, \log \frac{1}{\delta})$.*

Remark 1. We intentionally present the runtime at a high level (polynomial time) because our reduction is not aimed at getting the optimal approximation ratio, not the tightest possible runtime bound. The major bottleneck is that we call $\mathcal{A}$ with $\tilde{O}\left(\frac{M^2}{\epsilon^2}\right)$ copies of the decision variables. Below, we also show how a simple modification of the algorithm (operating on only n decision variables) obtains approximation ratio $\alpha(1 - \frac{1}{e})$.

Remark 2. The RASCAL algorithm that we introduce for the continuous setting provides an algorithm $\mathcal{A}$ which may be used in this reduction with $\alpha = \left(1 - \frac{1}{e}\right)$. Hence, we immediately obtain a $\left(1 - \frac{1}{e}\right)$-approximate algorithm for the discrete portfolio optimization problem.

Proof. Define $OPT = \max_{p \in \Delta(\mathcal{I})} \text{CVaR}_\alpha \left(\mathbb{E}_{S \sim p} \left[f(S, y) \right] \right)$ and p be the maximizing distribution. We first claim

Lemma 8. *There is a distribution q with support on at most r sets which satisfies*

$$CVaR_\alpha\left(\mathop{\mathbb{E}}_{S\sim q}[f(S,y)]\right) \geq OPT - \epsilon$$

Proof. A stronger result is established in Wilder et al. [5]. Their Lemma 4 shows that there is a uniform distribution q on r sets which satisfies $\mathbb{E}_{S\sim q}[f(S,y)] \geq \mathbb{E}_{S\sim p}[f(S,y)] - \epsilon$ for all y. It is easy to check that since q has value ϵ-close to p for each y individually, it is also ϵ-close in aggregate CVaR value. □

In what follows, we will let q be such a distribution. Since optimizing over such bounded-support mixture distributions is computationally difficult, we now expand our search to a convex set which contains the promised good distribution q. Specifically, we optimize over the set of all distributions which are mixtures of r product distributions. We will maintain r decision variables $\boldsymbol{x}^1 ... \boldsymbol{x}^r \in \times_{i=1}^r \mathcal{P}$. We will consider the problem of maximizing $\text{CVaR}_\alpha\left(\frac{1}{r}\sum_{i=1}^r \mathbb{E}_{S\sim p_{\boldsymbol{x}^i}}\left[f(S,y)\right]\right)$. This is the CVaR of the mixture distribution on the r product distributions $\boldsymbol{x}^1 ... \boldsymbol{x}^r$. Note that the distribution q is an instance of such a distribution where $\boldsymbol{x}^i$ is the indicator vector for the set S_i (i.e., a product distribution which deterministically returns S_i) . Hence, by the above guarantee for q we have:

$$\max_{\boldsymbol{x}^1 ... \boldsymbol{x}^r \in \times_{i=1}^r \mathcal{P}} \text{CVaR}_\alpha\left(\frac{1}{r}\sum_{i=1}^r \mathop{\mathbb{E}}_{S\sim p_{\boldsymbol{x}^i}}\left[f(S,y)\right]\right) \geq \text{CVaR}_\alpha\left(\mathop{\mathbb{E}}_{S\sim q}\left[f(S,y)\right]\right) \geq OPT - \epsilon.$$

Thus, we can solve the first maximization problem to obtain a solution with value ϵ-close to the optimum. Here we use the multilinear extension. Note that in the optimization problem above, $\mathbb{E}_{S\sim p_{\boldsymbol{x}^i}}\left[f(S,y)\right]$ is exactly $F(\boldsymbol{x}^i, y)$, and $\frac{1}{r}\sum_{i=1}^r F(\boldsymbol{x}^i, y)$ is DR-submodular since it is a convex combination of DR-submodular functions. Hence, we can obtain a solution to the optimization problem with value at least $\alpha OPT - \epsilon$ by applying $\mathcal{A}$, the promised algorithm for continuous submodular CVaR optimization (line 3 of PortfolioCVaR). The only remaining issue is to convert the resulting mixture of product distributions into a distribution over elements of $\mathcal{I}$ with equivalent solution quality. Here, we use the swap rounding procedure. Via Lemma 7, it suffices to draw $O\left(\frac{\log\frac{sr}{\delta}}{\epsilon^3}\right)$ samples for each of the $\boldsymbol{x}^i$. Let the uniform distribution on the combined set of samples be q'. Via union bound over the r different product distributions, we have that $\mathbb{E}_{S\sim q'}[f(S,y)] \geq \frac{1}{r}\sum_{i=1}^r F(\boldsymbol{x}^i) - \epsilon$ holds for each $y \in \mathcal{Y}$ with probability at least $1-\delta$. Conditioning on this event, we have that

$$\begin{aligned}\text{CVaR}_\alpha\left(\mathop{\mathbb{E}}_{S\sim q'}[f(S,y)]\right) &\geq \text{CVaR}_\alpha\left(\frac{1}{r}\sum_{i=1}^r F(\boldsymbol{x}^i, y)\right) - \epsilon \\ &\geq \alpha OPT - 2\epsilon\end{aligned}$$

and now it suffices to apply the above argument with $\epsilon' = \frac{\epsilon}{2}$ to obtain the theorem. □

We close by noting that it is possible to obtain a more computationally efficient algorithm by maintaining only a single copy of the decision variables ($r = 1$), at the cost of a factor $(1-\frac{1}{e})$ in the approximation ratio:

Theorem 2. *Setting $r = 1$,* PortfolioCVaR *returns a solution with value at least $\alpha\left(1-\frac{1}{e}\right)OPT - \epsilon$ for any discrete portfolio optimization problem.*

Proof. The reason that we needed to introduce many copies is to guarantee that our feasible set includes a distribution which is ϵ-close in CVaR value to the optimal distribution p. However, we can use a known result for submodular functions to guarantee that the feasible set includes a $(1-\frac{1}{e})$-approximate solution even if we take $r=1$. Specifically, we use the *correlation gap* result of Agrawal et al. [1]. Let $D(\boldsymbol{x})$ be the set of all distributions with marginals $\boldsymbol{x}$. For any submodular function f, Agrawal et al. [1] prove that

$$\max_{\boldsymbol{x}\in\mathcal{P}, q\in D(\boldsymbol{x})} \frac{\mathbb{E}_{S\sim q}[f(S)]}{\mathbb{E}_{S\sim p_{\boldsymbol{x}}}[f(S)]} \leq \frac{e}{e-1}$$

Hence, we can guarantee that

$$\max_{\boldsymbol{x}\in\mathcal{P}} \mathrm{CVaR}_\alpha \left(\mathop{\mathbb{E}}_{S\sim p_{\boldsymbol{x}}} [f(S,y)] \right) \geq \left(1-\frac{1}{e}\right) OPT$$

and from here, the same argument as before shows that applying PortfolioCVaR with $r=1$ results in a distribution with value at least $\alpha\left(1-\frac{1}{e}\right) OPT - \epsilon$. ☐

References

[1] Shipra Agrawal, Yichuan Ding, Amin Saberi, and Yinyu Ye. Correlation robust stochastic optimization. In *SODA*, 2010.

[2] Gruia Calinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM J. Comput.*, 40(6):1740–1766, 2011.

[3] Shaddin Dughmi and Haifeng Xu. Algorithmic persuasion with no externalities. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 351–368. ACM, 2017.

[4] Rishabh K. Iyer, Stefanie Jegelka, and Jeff A. Bilmes. Monotone closure of relaxed constraints in submodular optimization: Connections between minimization and maximization. In *UAI*, 2014.

[5] Bryan Wilder. Equilibrium computation and robust optimization in zero sum games with submodular structure. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, 2018.