Constrained Submodular Maximization: Beyond $1/e$

Alina Ene
Department of Computer Science
Boston University
Boston, USA
Email: aene@bu.edu

Huy L. Nguyễn
College of Computer and Information Science
Northeastern University
Boston, USA
Email: hlnguyen@cs.princeton.edu

Abstract—**In this work, we present a new algorithm for maximizing a non-monotone submodular function subject to a general constraint. Our algorithm finds an approximate fractional solution for maximizing the multilinear extension of the function over a down-closed polytope. The approximation guarantee is** 0.372 **and it is the first improvement over the** $1/e$ **approximation achieved by the unified Continuous Greedy algorithm [Feldman *et al.*, FOCS 2011].**

Keywords-**submodular functions; maximization;**

I. INTRODUCTION

A set function $f : 2^V \to \mathbb{R}$ is *submodular* if for every $A, B \subseteq V$, we have $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$. Submodular functions naturally arise in a variety of contexts, both in theory and practice. From a theoretical perspective, submodular functions provide a common abstraction of cut functions of graphs and digraphs, Shannon entropy, weighted coverage functions, and log-determinants. From a practical perspective, submodular functions are used in a wide range of application domains from machine learning to economics. In machine learning, it is used for document summarization [15], sensor placement [12], exemplar clustering [9], potential functions for image segmentation [10], etc. In an economics context, it can be used to model market expansion [6], influence in social networks [11], etc. The core mathematical problem underpinning many of these applications is the meta problem of maximizing a submodular objective function subject to some constraints, i.e., $\max_{S \in \mathcal{S}} f(S)$ where $\mathcal{S}$ is a down-closed family of sets[1].

A common approach to this problem is a two-step framework based on the multilinear extension F of f, a continuous function that extends f to the domain $[0, 1]^V$. The program first (1) maximizes $F(\mathbf{x})$ subject to $\mathbf{x} \in C$ where C is a convex relaxation of $\mathcal{S}$, and then (2) rounds $\mathbf{x}$ to an integral solution in $\mathcal{S}$. This paradigm has been very successful and it has led to the current best approximation algorithms for a wide variety of constraints including cardinality constraints, knapsack constraints, matroid constraints, etc. The contention resolution scheme framework of Chekuri *et al.* [5] gives

a generic way to perform step 2 of the program. Thus, many recent works [7], [4] focus on improving step 1, as it immediately leads to improved algorithms when combined with the known rounding schemes. Feldman *et al.* [7] give a beautiful generalization of the continuous greedy algorithm [18], achieving an $1/e$ approximation for step 1 for any down-closed and solvable polytope[2] C. While it is known that the continuous greedy algorithm is optimal when we restrict to monotone functions[3], it is not known if the $1/e$ approximation of the extension by [7] is optimal. Recently, the work of Buchbinder *et al.* [1] shows that, for the special case of a *cardinality* constraint, it is possible to beat $1/e$. However, this result still leaves open the possibility that an $1/e$ approximation is best possible for a harder constraint. This possibility is consistent with our current knowledge: the best known hardness for a cardinality constraint is 0.491 while the best known hardness for a matroid constraint is 0.478 [8], suggesting that the matroid constraint may be strictly harder to approximate. In this paper, we rule out this possibility and show that it is possible to go beyond the $1/e$ barrier in the same generic setting as considered by [7].

Theorem 1. *Let f be a non-negative, non-monotone, submodular function f. Let F be the multilinear extension of f. Let C be a down-closed and solvable polytope. There is an efficient algorithm that constructs a solution $\mathbf{x} \in C$ such that $F(\mathbf{x}) \geq 0.372 \cdot F(\text{OPT})$, where OPT is an optimal integral solution to the problem $\max_{\mathbf{x} \in C \cap \{0,1\}^V} F(\mathbf{x})$.*

In order to keep the analysis as simple as possible, we have not tried to optimize the constant in the theorem above. We believe a better constant can be obtained using the techniques in this paper, and we hope that our work will lead to improved approximation guarantees for the problem.

Using known rounding techniques, we obtain improved approximation guarantees for several classes of constraints. In particular, we obtain a $0.372 - o(1)$ approximation for submodular maximization subject to a matroid constraint and a $0.372 - \epsilon$ approximation for a constant number of knapsack

---

[1]A family of sets $\mathcal{S}$ is down-closed if for all $A \subset B$, if $B \in \mathcal{S}$ then $A \in \mathcal{S}$.

[2]A polytope C is solvable if there is an oracle for optimizing linear functions over C, i.e., for solving $\max_{\mathbf{c} \in C} \langle \mathbf{v}, \mathbf{c} \rangle$ for any vector $\mathbf{v}$.

[3]A function $f : 2^V \to \mathbb{R}$ is monotone if $f(A) \leq f(B)$ for all $A \subseteq B$.

IEEE
computer
society

constraints, which improve over the $1/e - o(1)$ and $1/e - \epsilon$ approximations [7].

A. Our techniques

Our starting point is the unified continuous greedy algorithm [7] achieving the approximation factor $1/e$. The algorithm grows a solution $\mathbf{x}$ over the time interval $[0, 1]$, improving $F(\mathbf{x})$ in each time step by an amount proportional to $(1 - \|\mathbf{x}\|_\infty)F(\text{OPT})$. As $\mathbf{x}$ increases over time, the gain starts out large and decreases to 0 at the end of the process. The change in $\mathbf{x}$ in each time step is proportional[4] to $\mathbf{v} = \text{argmax}_{\mathbf{c} \in C}\langle \nabla F(\mathbf{x}) \circ (\mathbf{1} - \mathbf{x}), \mathbf{c}\rangle$. Notice that we can improve the gain $((1 - \|\mathbf{x}\|_\infty)F(\text{OPT}))$ by slowing down the growth of $\|\mathbf{x}\|_\infty$. Thus in our algorithm, we add a new constraint $\|\mathbf{c}\|_\infty \leq \alpha$ and have $\mathbf{v} = \text{argmax}_{\mathbf{c} \in C, \|\mathbf{c}\|_\infty \leq \alpha}\langle \nabla F(\mathbf{x}) \circ (\mathbf{1} - \mathbf{x}), \mathbf{c}\rangle$. How does this change affect the performance of the algorithm and in particular, the quantity $\langle \nabla F(\mathbf{x}) \circ (\mathbf{1} - \mathbf{x}), \mathbf{v}\rangle$? Intuitively, if there is a second solution other than OPT with value close to OPT, we can pick $\mathbf{v}$ to be a mixture of OPT and this solution and have a new solution comparable to OPT but with lower ℓ_∞ norm. Thus, if this is the case, we do not lose very much in $\langle \nabla F(\mathbf{x}) \circ (\mathbf{1} - \mathbf{x}), \mathbf{v}\rangle$ while simultaneously increasing the gain $(1 - \|\mathbf{x}\|_\infty)F(\text{OPT})$. On the other hand, if there is no such solution, the crucial insight is that $\mathbf{v}$ must be *well correlated* with OPT. Thus we can identify a good fraction of OPT by searching for $\text{argmax}_{\mathbf{x} \leq \mathbf{v}} F(\mathbf{x})$. (Here we crucially use the fact that C is a down-closed polytope.) This problem turns out to be not very different from the *unconstrained* setting and we can use a variant of the double greedy algorithm of Buchbinder *et al.* [2] to find a good solution.

The above description is an intuitive but simplified overview of the algorithm. Nevertheless, each routine mentioned above corresponds to a part of the algorithm described in Section IV. The main technical difficulty is in formalizing the statement "$\mathbf{v}$ must be well correlated with OPT" and the subsequent identification of (a large part of) OPT. We describe how to overcome these difficulties in Section V-B.

Comparison with the work of [1] for a cardinality constraint. The idea of adding a new constraint $\|\mathbf{c}\|_\infty \leq \alpha$ is inspired by a recent algorithm by [1] for the cardinality constraint. In their setting, the goal is to maximize $F(\mathbf{x})$ subject to a constraint $\|\mathbf{x}\|_1 \leq k$. In the ith step, their algorithm picks $2(k - i)$ elements with maximum marginal gain and randomly adds one of them to the current solution. This can be viewed as an analog of adding a constraint $\|\mathbf{c}\|_\infty \leq k/(2k - 2i)$ when the time is in the range $[(i - 1)/k, i/k]$. This more sophisticated use of varying thresholds depending on the time is allowed in our framework

but in the simple solution presented here, we just use a fixed threshold throughout.

When the marginal gain ($\langle \nabla F(\mathbf{x}) \circ (1 - \mathbf{x}), \mathbf{v}\rangle$ in our case) is small, they also use a variant of the double greedy algorithm to finish the solution. From the $2(k - i)$ elements that are picked in the last iteration, their algorithm picks $k - i$ elements using a sophisticated variant of double Greedy, and add them to the i elements it picked in the previous iterations. Unfortunately this step crucially uses the structure of the solution when the constraint is a cardinality constraint. From the point of view of a continuous Greedy algorithm, this is analogous to doubling a fractional solution and then selecting half of the coordinates. It is not clear what the analog should be in the general case. Instead, we compute a different solution $\mathbf{p} = \text{argmax}_{\mathbf{c} \in C}\langle \nabla F(\mathbf{x}) \circ (1 - \mathbf{x}), \mathbf{c}\rangle$ (dropping the dampening constraint) and use standard double greedy to find $\mathbf{z} = \text{argmax}_{\mathbf{c} \leq \mathbf{p}} F(\mathbf{c})$. While $\mathbf{v}$ and $\mathbf{p}$ could potentially be very different, we manage to connect $\langle \nabla F(\mathbf{x}) \circ (1 - \mathbf{x}), \mathbf{v}\rangle$ and $\langle \nabla F(\mathbf{x}) \circ (1 - \mathbf{x}), \mathbf{p}\rangle$, which is enough to prove the existence of $\mathbf{z} \leq \mathbf{p}$ with large $F(\mathbf{z})$.

In summary, in this work, we introduce several novel insights that lead to a much more general algorithm and analysis. We believe that our algorithm and analysis are conceptually much simpler and cleaner, and we hope that our techniques will lead to further improvements for the problem.

B. Related work

Submodular maximization problems are well-studied and it is hard to do justice to all previous literature on this subject. Starting with the seminal work Nemhauser et al. [16], the classical approaches to these problems are largely combinatorial and based on greedy and local search algorithms. Over the years, with increasing sophistication, this direction has led to many tight results such as the algorithm of [17] for a knapsack constraint, and the current best results for constraints such as multiple matroid constraints [14]. In the last few years, another approach emerged [3] that follows the popular paradigm in approximation algorithms of optimizing a continuous relaxation and rounding the resulting fractional solution. A key difficulty that separates the submodular setting from the classical setting is that even finding a *fractional solution* may be quite challenging, and in particular it is **NP**-hard to solve the continuous relaxation for maximizing submodular functions that is based on the multilinear extension. Thus, a line of work has been developed to *approximately* optimize this relaxation [3], [13], [8], [7] culminating in the work [7], which we extend here.

II. PRELIMINARIES

Here we consider the problem of maximizing a submodular function f subject to a downward closed convex constraint $\mathbf{x} \in C$.

---

[4]We use $\mathbf{x} \circ \mathbf{y}$ to denote the vector whose i-th coordinate is $\mathbf{x}_i \cdot \mathbf{y}_i$.

We use the following notation. Let $n = |V|$. We write $\mathbf{x} \leq \mathbf{y}$ if $\mathbf{x}_i \leq \mathbf{y}_i$ for all coordinates $i \in [n]$. Let $\mathbf{x} \circ \mathbf{y}$ denote the vector whose i-th coordinate is $\mathbf{x}_i \cdot \mathbf{y}_i$. Let $\mathbf{x} \vee \mathbf{y}$ (resp. $\mathbf{x} \wedge \mathbf{y}$) be the vector whose i-th coordinate is $\max\{\mathbf{x}_i, \mathbf{y}_i\}$ (resp. $\min\{\mathbf{x}_i, \mathbf{y}_i\}$). Let $\mathbf{1}_S \in \{0,1\}^V$ denote the indicator vector of $S \subseteq V$, i.e., the vector that has a 1 in entry i if and only if $i \in S$.

Let $F : [0,1]^n \to \mathbb{R}_+$ denote the multilinear extension of f:

$$F(\mathbf{x}) = \mathbb{E}[f(R(\mathbf{x}))] = \sum_{S \subseteq V} f(S) \prod_{i \in S} \mathbf{x}_i \prod_{i \in V \setminus S} (1 - \mathbf{x}_i),$$

where $R(\mathbf{x})$ is a random subset of V where each $i \in V$ is included independently at random with probability $\mathbf{x}_i$. We use ∇F to denote the gradient of F and we use $\frac{\partial F}{\partial \mathbf{x}_i}$ to denote the i-th coordinate of the gradient of F.

The multilinear extension has the following well-known properties, see for instance [18].

Claim 2. $\frac{\partial F}{\partial x_i}(\mathbf{x}) = F(\mathbf{x} \vee \mathbf{1}_{\{i\}}) - F(\mathbf{x} \wedge \mathbf{1}_{V \setminus \{i\}})$.

Proof: Note that

$$F(\mathbf{x}) = \mathbf{x}_i \cdot F(\mathbf{x} \vee \mathbf{1}_{\{i\}}) + (1 - \mathbf{x}_i) \cdot F(\mathbf{x} \wedge \mathbf{1}_{V \setminus \{i\}}).$$

Thus if we take the partial derivative with respect to $\mathbf{x}_i$, we obtain the claim. ∎

Claim 3. *If $\mathbf{x} \leq \mathbf{y}$ then $\nabla F(\mathbf{x}) \geq \nabla F(\mathbf{y})$.*

Proof: Fix a coordinate i. Since $\mathbf{x} \leq \mathbf{y}$, submodularity implies that

$$F(\mathbf{x} \vee \mathbf{1}_{\{i\}}) - F(\mathbf{x} \wedge \mathbf{1}_{V \setminus \{i\}}) \geq F(\mathbf{y} \vee \mathbf{1}_{\{i\}}) - F(\mathbf{y} \wedge \mathbf{1}_{V \setminus \{i\}}).$$

By Claim 2, the left-hand side of the inequality above is $\frac{\partial F}{\partial x_i}(\mathbf{x})$ and the right-hand side is $\frac{\partial F}{\partial x_i}(\mathbf{y})$. ∎

Claim 4. *F is concave along any line of direction $\mathbf{d} \geq \mathbf{0}$. That is, for any vector $\mathbf{x}$, the function $\phi : \mathbb{R} \to \mathbb{R}$ such that $\phi(t) = F(\mathbf{x} + t \cdot \mathbf{d})$ for every $t \in \mathbb{R}$ for which $\mathbf{x} + t \cdot \mathbf{d} \in [0,1]^V$ is concave.*

Proof sketch: By submodularity, we can verify that $\phi''(t) \leq 0$ and thus ϕ is concave. □

III. DOUBLE GREEDY ALGORITHM FOR A BOX CONSTRAINT

In this section, we describe an algorithm for maximizing the multilinear extension subject to a box constraint: given $\mathbf{u}$ and $\mathbf{v}$, find $\max_{\mathbf{u} \leq \mathbf{x} \leq \mathbf{v}} F(\mathbf{x})$. The algorithm is similar to the Double Greedy algorithm of [2] and it is given in Figure 1.

The proof of the following lemma is similar to the analysis of the Double Greedy algorithm and it can be found in Appendix A.

```
 1:  u^(0) = u, v^(0) = v
 2:  for i ∈ [n] do
 3:      a_i ← F(u^(i-1) ∨ 1_{i}) − F(u^(i-1) ∧ 1_{V\{i}})
 4:      b_i ← F(v^(i-1) ∧ 1_{V\{i}}) − F(v^(i-1) ∨ 1_{i})
 5:      a'_i = max(a_i, 0), b'_i = max(b_i, 0)
 6:      if a'_i + b'_i ≠ 0 then
 7:          u^(i) ← u^(i-1) + (a'_i/(a'_i+b'_i)) · (v_i − u_i) · 1_{i}
 8:          v^(i) ← v^(i-1) − (b'_i/(a'_i+b'_i)) · (v_i − u_i) · 1_{i}
 9:      else
        ▷ Update u_i^(i) and v_i^(i) to any common value
        ▷ We set u_i^(i) = v_i^(i) = v_i^(i-1)
10:         u^(i) ← u^(i-1) + (v_i − u_i) · 1_{i}
11:         v^(i) ← v^(i-1)
12:     end if
13: end for
14: return u^(n)
```

Figure 1. Double Greedy algorithm for a box constraint $\{\mathbf{x} \colon \mathbf{u} \leq \mathbf{x} \leq \mathbf{v}\}$.

Lemma 5. *Algorithm 0 finds a solution $\mathbf{x}$ to the problem $\max_{\mathbf{u} \leq \mathbf{x} \leq \mathbf{v}} F(\mathbf{x})$ such that*

$$F(\mathbf{x}) \geq \frac{1}{2} \cdot F(\text{OPT}) + \frac{1}{4} \cdot F(\mathbf{u}) + \frac{1}{4} \cdot F(\mathbf{v}),$$

where OPT *is an optimal solution to the problem* $\max_{\mathbf{u} \leq \mathbf{x} \leq \mathbf{v}} F(\mathbf{x})$.

Alternatively, we can reduce the problem with a box constraint to the unconstrained problem by defining a suitable submodular function $g(S) = F(\mathbf{u} + (\mathbf{v} - \mathbf{u}) \circ \mathbf{1}_S)$. We can show that this reduction is correct using the same argument given in the appendix.

IV. THE ALGORITHM

In this section, we describe our main algorithm for the problem $\max_{\mathbf{x} \in C} F(\mathbf{x})$. We first give a continuous version of our algorithm; in order to efficiently implement the algorithm, we discretize it using a standard approach, and we give the details in Appendix A.

The algorithm picks the best out of two solutions. The first solution is constructed by running the Continuous Greedy algorithm *with an additional dampening constraint*[5] as long as the marginal gain remains high despite the dampening constraint, and then finishing the solution via the standard Continuous Greedy algorithm for the remaining time. The second solution is constructed by running Double Greedy exactly when the marginal gain becomes low because of the dampening constraint, which must happen early if the first solution is not good. Since we do not know precisely

---
[5]The dampening constraint imposes an ℓ_∞ constraint in line 5 of Figure 2

```
 1: Initialize $\mathbf{x}^* = \mathbf{0}$
 2: for $\theta \in [0, 1]$ do
      ($\triangleright$) Dampened Continuous Greedy stage
 3:     Initialize $\mathbf{x}^{(0)} = \mathbf{0}$
 4:     for $t \in [0, \theta]$ do
 5:         $\mathbf{v}^{(t)} = \mathrm{argmax}_{\mathbf{c} \in C, \|\mathbf{c}\|_\infty \leq \alpha} \langle \nabla F(\mathbf{x}^{(t)}) \circ (\mathbf{1} - \mathbf{x}^{(t)}), \mathbf{c} \rangle$
 6:         Update $\mathbf{x}^{(t)}$ according to $\frac{d\mathbf{x}^{(t)}}{dt} = \mathbf{v}^{(t)} \circ (\mathbf{1} - \mathbf{x}^{(t)})$
 7:     end for

      ($\triangleright$) Standard Continuous Greedy stage
 8:     Initialize $\mathbf{y}^{(\theta)} = \mathbf{x}^{(\theta)}$
 9:     for $t \in (\theta, 1]$ do
10:         $\mathbf{v}^{(t)} = \mathrm{argmax}_{\mathbf{c} \in C} \langle \nabla F(\mathbf{y}^{(t)}) \circ (\mathbf{1} - \mathbf{y}^{(t)}), \mathbf{c} \rangle$
11:         Update $\mathbf{y}^{(t)}$ according to $\frac{d\mathbf{y}^{(t)}}{dt} = \mathbf{v}^{(t)} \circ (\mathbf{1} - \mathbf{y}^{(t)})$
12:     end for

      ($\triangleright$) Double Greedy stage
13:     $\mathbf{p} = \mathrm{argmax}_{\mathbf{c} \in C} \langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), \mathbf{c} \rangle$
14:     Find $\mathbf{z}$ approximating $\mathrm{argmax}_{\mathbf{c} \leq \mathbf{p}} F(\mathbf{c})$ using the Double Greedy algorithm (see Figure 1)

      ($\triangleright$) Update the best solution
15:     if $F(\mathbf{y}^{(1)}) > F(\mathbf{x}^*)$ then
16:         $\mathbf{x}^* = \mathbf{y}^{(1)}$
17:     end if
18:     if $F(\mathbf{z}) > F(\mathbf{x}^*)$ then
19:         $\mathbf{x}^* = \mathbf{z}$
20:     end if
21: end for
22: return $\mathbf{x}^*$
```

Figure 2.   Continuous algorithm for a general constraint

when the marginal gain becomes low, the algorithm tries all possible values via the outer for loop on line 2.

V. The analysis

In this section, we analyze the algorithm and show that it achieves an approximation greater than 0.372. We remark that the analysis in Subsection V-A is fairly standard, and the crux of the analysis is in Subsection V-B.

In the following, we fix an iteration θ of the outer loop and we analyze the solutions constructed in that iteration. In the remainder of this section, all of the vectors $\mathbf{x}^{(\cdot)}$, $\mathbf{y}^{(\cdot)}$, etc. refer to the vectors during iteration θ.

A. Analysis of the solution $\mathbf{y}^{(1)}$

In this section, we analyze the solution $\mathbf{y}^{(1)}$ constructed in any given iteration $\theta \in [0, 1]$.

Theorem 6. *Let $\theta \in [0, 1]$ and let $\mathbf{y}^{(1)}$ be the solution constructed by the Algorithm in Figure 2 in iteration θ of the outer loop. We have*

$$F(\mathbf{y}^{(1)}) \geq \frac{1}{e} \left(e^\theta F(\mathbf{x}^{(\theta)}) + (1 - \theta) e^{(1-\alpha)\theta} F(\mathrm{OPT}) \right).$$

We devote the rest of this section to the proof of Theorem 6. In the remainder of this section, all of the vectors $\mathbf{x}^{(\cdot)}$, $\mathbf{y}^{(\cdot)}$, etc. refer to the vectors during iteration θ.

We start by upper bounding the ℓ_∞ norm of $\mathbf{x}^{(t)}$ and $\mathbf{y}^{(t)}$.

Lemma 7. *Consider the following process. Let time run from $t = t_0 \geq 0$ to $t = t_1 \leq 1$. Let $\mathbf{u}^{(t)}$ be a vector updated according to*

$$\frac{d\mathbf{u}^{(t)}}{dt} = \mathbf{v}^{(t)} \circ (\mathbf{1} - \mathbf{u}^{(t)}),$$

where $\mathbf{v}^{(t)}$ is a vector such that $\|\mathbf{v}^{(t)}\|_\infty \leq \delta$. Then $\|\mathbf{u}^{(t)}\|_\infty \leq 1 + (\|\mathbf{u}^{(t_0)}\|_\infty - 1)e^{-\delta(t-t_0)}$ for each $t \in [t_0, t_1]$.

Proof: The i-th coordinate $\mathbf{u}_i^{(t)}$ of $\mathbf{u}^{(t)}$ is updated according to

$$\frac{d\mathbf{u}_i^{(t)}}{dt} = \mathbf{v}_i^{(t)}(1 - \mathbf{u}_i^{(t)}) \leq \delta(1 - \mathbf{u}_i^{(t)}).$$

By solving the differential inequality above, we obtain $\mathbf{u}_i^{(t)} \leq 1 + C \cdot e^{-\delta t}$, where C is a constant. Using the initial condition $1 + C \cdot e^{-\delta t_0} = \mathbf{u}_i^{(t_0)}$, we obtain $C = (\mathbf{u}_i^{(t_0)} - 1)e^{\delta t_0}$ and thus $\mathbf{u}_i^{(t)} \leq 1 + (\mathbf{u}_i^{(t_0)} - 1)e^{-\delta(t-t_0)}$. ∎

Corollary 8. *For every $t \in [0, \theta]$, $\|\mathbf{x}^{(t)}\|_\infty \leq 1 - e^{-\alpha t}$.*

Proof: The vector $\mathbf{x}^{(t)}$ starts at $\mathbf{0}$ and it is updated according to the update rule in line 6, where $\mathbf{v}^{(t)}$ is a vector of ℓ_∞ norm at most α. Thus it follows from Lemma 7 that $\|\mathbf{x}^{(t)}\|_\infty \leq 1 - e^{-\alpha t}$. ∎

Corollary 9. *For every $t \in [\theta, 1]$, $\|\mathbf{y}^{(t)}\|_\infty \leq 1 - e^{(1-\alpha)\theta - t}$.*

Proof: The vector $\mathbf{y}^{(t)}$ starts at $\mathbf{y}^{(\theta)} = \mathbf{x}^{(\theta)}$ and it is updated according to the update rule in line 11, where $\mathbf{v}^{(t)}$ is a vector of ℓ_∞ norm at most 1. Thus it follows from Lemma 7 and the upper bound on $\|\mathbf{x}^{(\theta)}\|_\infty$ given by Corollary 8 that

$$\|\mathbf{y}^{(t)}\|_\infty \leq 1 + (\|\mathbf{x}^{(\theta)}\|_\infty - 1)e^{-(t-\theta)} \leq 1 - e^{(1-\alpha)\theta - t}.$$
∎

We will also need the following lemma that was shown in [7].

Lemma 10 ([7, Lemma III.5]). *Let $\mathbf{x} \in [0, 1]^n$ and let $S \subseteq V$. We have*

$$F(\mathbf{x} \vee \mathbf{1}_S) \geq (1 - \|\mathbf{x}\|_\infty)f(S).$$

Proof of Theorem 6: Using the chain rule, for every $t \in [\theta, 1]$, we have

$$
\begin{aligned}
\frac{dF(\mathbf{y}^{(t)})}{dt} &= \left\langle \nabla F(\mathbf{y}^{(t)}), \frac{d\mathbf{y}^{(t)}}{dt} \right\rangle \\
&= \left\langle \nabla F(\mathbf{y}^{(t)}), \mathbf{v}^{(t)} \circ (\mathbf{1} - \mathbf{y}^{(t)}) \right\rangle \\
&= \left\langle \nabla F(\mathbf{y}^{(t)}) \circ (\mathbf{1} - \mathbf{y}^{(t)}), \mathbf{v}^{(t)} \right\rangle \\
&\geq \left\langle \nabla F(\mathbf{y}^{(t)}) \circ (\mathbf{1} - \mathbf{y}^{(t)}), \mathrm{OPT} \right\rangle \\
&\geq F(\mathbf{y}^{(t)} \vee \mathrm{OPT}) - F(\mathbf{y}^{(t)}) \\
&\geq e^{(1-\alpha)\theta - t} \cdot F(\mathrm{OPT}) - F(\mathbf{y}^{(t)}).
\end{aligned}
$$

The last line above follows from Corollary 9 and Lemma 10. By solving the differential inequality, we obtain

$$
F(\mathbf{y}^{(t)}) \geq \frac{1}{e^t} \left(e^\theta F(\mathbf{y}^{(\theta)}) + (t - \theta)e^{(1-\alpha)\theta} F(\mathrm{OPT}) \right),
$$

and thus

$$
F(\mathbf{y}^{(1)}) \geq \frac{1}{e} \left(e^\theta F(\mathbf{y}^{(\theta)}) + (1 - \theta)e^{(1-\alpha)\theta} F(\mathrm{OPT}) \right).
$$

The theorem now follows, since $\mathbf{y}^{(\theta)} = \mathbf{x}^{(\theta)}$. $\qquad\square$

B. Analysis of the solution $\mathbf{z}$

In this section, we analyze the solution $\mathbf{z}$ constructed using the Double Greedy algorithm, and this is the crux of our argument.

Theorem 11. *Let* $\theta \in [0, 1]$ *and let* $\mathbf{z}$ *be the solution constructed by the algorithm in Figure 2 in iteration* θ *of the outer loop. If* $\alpha \geq 1/2$*, we have*

$$
F(\mathbf{z}) \geq \frac{1}{2(1-\alpha)}(e^{-\alpha\theta}F(\mathrm{OPT}) - F(\mathbf{x}^{(\theta)})-
$$
$$
\langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), \mathbf{v}^{(\theta)} \rangle).
$$

We first give some intuition for the approach. Consider the solution $\mathbf{x}$ constructed by the *dampened* Continuous Greedy algorithm. The rate of growth of $F(\mathbf{x})$ is given by the inner product $\langle \nabla F(\mathbf{x}) \circ (\mathbf{1} - \mathbf{x}), \mathbf{v} \rangle$ and thus the crux of the analysis is to understand how this inner product evolves over time. The inner product $\langle \nabla F(\mathbf{x}) \circ (\mathbf{1} - \mathbf{x}), \mathbf{v} \rangle$ is always at least $\langle \nabla F(\mathbf{x}) \circ (\mathbf{1} - \mathbf{x}), \alpha\mathrm{OPT} \rangle$ and intuitively we should gain proportional to the difference between the two. If this difference is very small, the key insight is that once we drop the dampening constraint and compute the vector that maximizes $\langle \nabla F(\mathbf{x}) \circ (\mathbf{1} - \mathbf{x}), \mathbf{c} \rangle$ over all feasible vectors $\mathbf{c}$, we obtain a vector $\mathbf{p}$ that is well-correlated with OPT in the sense that $\mathbf{p} \wedge \mathrm{OPT}$ is a good solution. We formalize this intuition in the remainder of this section.

In the remainder of this section, we analyze the solution

$$
\widehat{\mathbf{z}} := (\mathbf{1} - \mathbf{x}^{(\theta)}) \circ (\mathbf{p} \wedge \mathrm{OPT}).
$$

By submodularity, we have

$$
F(\widehat{\mathbf{z}}) - F(\mathbf{0}) \geq F(\widehat{\mathbf{z}} + \mathbf{x}^{(\theta)}) - F(\mathbf{x}^{(\theta)}),
$$

and thus it suffices to analyze $F(\widehat{\mathbf{z}} + \mathbf{x}^{(\theta)})$.

Note that $\mathbf{x}^{(\theta)} \leq \widehat{\mathbf{z}} + \mathbf{x}^{(\theta)} \leq \mathbf{x}^{(\theta)} + (\mathbf{1} - \mathbf{x}^{(\theta)}) \circ \mathrm{OPT} = \mathbf{x}^{(\theta)} \vee \mathrm{OPT}$. Thus Claim 4 and Claim 3 give

$$
\begin{aligned}
&F(\mathbf{x}^{(\theta)} \vee \mathrm{OPT}) - F(\widehat{\mathbf{z}} + \mathbf{x}^{(\theta)}) \\
&\leq \langle \nabla F(\widehat{\mathbf{z}} + \mathbf{x}^{(\theta)}), (\mathbf{1} - \mathbf{x}^{(\theta)}) \circ ((\mathbf{1} - \mathbf{p}) \wedge \mathrm{OPT}) \rangle \\
&\leq \langle \nabla F(\mathbf{x}^{(\theta)}), (\mathbf{1} - \mathbf{x}^{(\theta)}) \circ ((\mathbf{1} - \mathbf{p}) \wedge \mathrm{OPT}) \rangle \\
&= \langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), (\mathbf{1} - \mathbf{p}) \wedge \mathrm{OPT} \rangle
\end{aligned}
$$

In the first inequality, we have used the fact that $(\mathbf{x}^{(\theta)} \vee \mathrm{OPT}) - (\widehat{\mathbf{z}} + \mathbf{x}^{(\theta)}) = (\mathbf{x}^{(\theta)} + (\mathbf{1} - \mathbf{x}^{(\theta)}) \circ \mathrm{OPT}) - (\widehat{\mathbf{z}} + \mathbf{x}^{(\theta)}) = (\mathbf{1} - \mathbf{x}^{(\theta)}) \circ ((\mathbf{1} - \mathbf{p}) \wedge \mathrm{OPT})$.

In order to upper bound $\langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), (\mathbf{1} - \mathbf{p}) \wedge \mathrm{OPT} \rangle$, we consider the following vector.

$$
\widehat{\mathbf{v}} := (1 - \alpha)((\mathbf{1} - \mathbf{p}) \wedge \mathrm{OPT}) + \alpha\mathbf{p}.
$$

The intuition behind the choice of $\widehat{\mathbf{v}}$ is to connect the fact that the inner product $\langle \nabla F(\mathbf{x}^{(\tau)}) \circ (\mathbf{1} - \mathbf{x}^{(\tau)}), \mathbf{v} \rangle$ is not much more than $\langle \nabla F(\mathbf{x}^{(\tau)}) \circ (\mathbf{1} - \mathbf{x}^{(\tau)}), \alpha\mathrm{OPT} \rangle$ to the insight that $\langle \nabla F(\mathbf{x}^{(\tau)}) \circ (\mathbf{1} - \mathbf{x}^{(\tau)}), (\mathbf{1} - \mathbf{p}) \wedge \mathrm{OPT} \rangle$ is small.

Since $\alpha \geq 1/2$, we have $1 - \alpha \leq \alpha$. Therefore, for each $i \in [n]$, we have

$$
\widehat{\mathbf{v}}_i = \begin{cases} (1 - \alpha)(1 - \mathbf{p}_i) + \alpha\mathbf{p}_i \leq \alpha & \text{if } i \in \mathrm{OPT} \\ \alpha\mathbf{p}_i \leq \alpha & \text{otherwise} \end{cases}
$$

Therefore $\|\widehat{\mathbf{v}}\|_\infty \leq \alpha$. Additionally, $\widehat{\mathbf{v}} \in C$, since it is a convex combination of two vectors in C (recall that C is downward closed and convex). It follows from the definition of $\mathbf{v}^{(\theta)}$ on line 5 that

$$
\begin{aligned}
&\langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), \mathbf{v}^{(\theta)} \rangle \\
&\geq \langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), \widehat{\mathbf{v}} \rangle \\
&= (1 - \alpha)\langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), (\mathbf{1} - \mathbf{p}) \wedge \mathrm{OPT} \rangle + \\
&\quad \alpha\langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), \mathbf{p} \rangle \\
&\geq (1 - \alpha)\langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), (\mathbf{1} - \mathbf{p}) \wedge \mathrm{OPT} \rangle + \\
&\quad \alpha\langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), \mathrm{OPT} \rangle \\
&\geq (1 - \alpha)\langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), (\mathbf{1} - \mathbf{p}) \wedge \mathrm{OPT} \rangle + \\
&\quad \alpha(F(\mathbf{x}^{(\theta)} \vee \mathrm{OPT}) - F(\mathbf{x}^{(\theta)}))
\end{aligned}
$$

By rearranging the inequality above, we obtain

$$
\begin{aligned}
&\langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), (\mathbf{1} - \mathbf{p}) \wedge \mathrm{OPT} \rangle \\
&\leq \frac{1}{1 - \alpha}\langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), \mathbf{v}^{(\theta)} \rangle - \\
&\quad \frac{\alpha}{1 - \alpha}\left(F(\mathbf{x}^{(\theta)} \vee \mathrm{OPT}) - F(\mathbf{x}^{(\theta)}) \right).
\end{aligned}
$$

By combining all of the inequalities, we obtain

$$\begin{aligned}
F(\widehat{\mathbf{z}}) &\geq F(\widehat{\mathbf{z}} + \mathbf{x}^{(\theta)}) - F(\mathbf{x}^{(\theta)})\\
&\geq F(\mathbf{x}^{(\theta)} \vee \text{OPT}) - F(\mathbf{x}^{(\theta)})-\\
&\quad \langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), (\mathbf{1} - \mathbf{p}) \wedge \text{OPT})\rangle\\
&\geq \frac{1}{1-\alpha}\Big(F(\mathbf{x}^{(\theta)} \vee \text{OPT}) - F(\mathbf{x}^{(\theta)})-\\
&\quad \langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), \mathbf{v}^{(\theta)}\rangle\Big)\\
&\geq \frac{1}{1-\alpha}\Big(e^{-\alpha\theta} F(\text{OPT}) - F(\mathbf{x}^{(\theta)})-\\
&\quad \langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), \mathbf{v}^{(\theta)}\rangle\Big)
\end{aligned}$$

In the last inequality, we have used Corollary 8 and Lemma 10 to lower bound $F(\mathbf{x}^{(\theta)} \vee \text{OPT})$ by $e^{-\alpha\theta}F(\text{OPT})$.

Since $\widehat{\mathbf{z}}$ is a candidate solution for the Double Greedy step on line 14, it follows from Lemma 5 that $F(\mathbf{z}) \geq \frac{1}{2}F(\widehat{\mathbf{z}})$, and the theorem follows.

C. Combining the two solutions

As we have shown above, in every iteration θ of the algorithm, we obtain two solutions $\mathbf{y}^{(1)}$ and $\mathbf{z}$ satisfying:

$$F(\mathbf{y}^{(1)}) \geq \frac{1}{e}\Big((1-\theta)e^{(1-\alpha)\theta} F(\text{OPT}) + e^{\theta} F(\mathbf{x}^{(\theta)})\Big)$$

$$\begin{aligned}
F(\mathbf{z}) \geq \frac{1}{2(1-\alpha)}\Big(&e^{-\alpha\theta} F(\text{OPT}) - F(\mathbf{x}^{(\theta)})-\\
&\langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), \mathbf{v}^{(\theta)}\rangle\Big)
\end{aligned}$$

In the following, we show that there is an iteration θ for which $\max\{F(\mathbf{y}^{(1)}), F(\mathbf{z})\} > C \cdot F(\text{OPT})$, where C is a constant that we will set later. We will proceed by contradiction and assume that $\max\{F(\mathbf{y}^{(1)}), F(\mathbf{z})\} \leq C \cdot F(\text{OPT})$ for all θ.

Note that the coefficient of $F(\mathbf{x}^{(\theta)})$ is positive in the first inequality above and it is negative in the second inequality, and there is a trade-off between the two solutions. We can get a handle on this trade-off as follows. To simplify matters, we take a convex combination of the two inequalities and eliminate $F(\mathbf{x}^{(\theta)})$. Thus we get that, for all $\theta \in [0,1]$,

$$\frac{1}{\frac{1}{2(1-\alpha)} + e^{\theta-1}}\Big(\frac{(2-\theta)e^{(1-\alpha)\theta-1}}{2(1-\alpha)} F(\text{OPT})-$$
$$\frac{e^{\theta-1}}{2(1-\alpha)}\langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), \mathbf{v}^{(\theta)}\rangle\Big) \leq C \cdot F(\text{OPT})$$

Thus, for all $\theta \in [0,1]$,

$$\begin{aligned}
&\langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), \mathbf{v}^{(\theta)}\rangle\\
&\geq \Big((2-\theta)e^{-\alpha\theta} - (e^{1-\theta} + 2(1-\alpha))C\Big)F(\text{OPT}).
\end{aligned}$$

Now note that

$$\begin{aligned}
&F(\mathbf{x}^{(\theta)})\\
&= \int_0^\theta \langle \nabla F(\mathbf{x}^{(t)}) \circ (\mathbf{1} - \mathbf{x}^{(t)}), \mathbf{v}^{(t)}\rangle)dt\\
&\geq \int_0^\theta \Big((2-t)e^{-\alpha t} - (e^{1-t} + 2(1-\alpha))C\Big)F(\text{OPT})dt\\
&= \left(\frac{e^{-\alpha t}(\alpha(t-2)+1)}{\alpha^2} - (-e^{1-t} + 2(1-\alpha)t)C\right)\Bigg|_{t=0}^{\theta}\\
&\quad \cdot F(\text{OPT})\\
&= \Bigg(\frac{e^{-\alpha\theta}(\alpha(\theta-2)+1) + 2\alpha - 1}{\alpha^2} -\\
&\quad (-e^{1-\theta} + 2(1-\alpha)\theta + e)C\Bigg)F(\text{OPT})
\end{aligned}$$

Therefore

$$\begin{aligned}
F(\mathbf{y}^{(1)}) \geq \Bigg(&(1-\theta)e^{(1-\alpha)\theta-1}\\
&+ \frac{e^{(1-\alpha)\theta-1}(\alpha(\theta-2)+1) + e^{\theta-1}(2\alpha-1)}{\alpha^2}\\
&- (-1 + 2(1-\alpha)\theta e^{\theta-1} + e^{\theta})C\Bigg)F(\text{OPT})
\end{aligned}$$

In order to obtain a contradiction, we need that the coefficient of $F(\text{OPT})$ in the inequality above is at least C for some $\theta \in [0,1]$ and some $\alpha \in [1/2, 1]$. Equivalently,

$$\begin{aligned}
&(1-\theta)e^{(1-\alpha)\theta-1}\\
&+ \frac{e^{(1-\alpha)\theta-1}(\alpha(\theta-2)+1) + e^{\theta-1}(2\alpha-1)}{\alpha^2}\\
&- (2(1-\alpha)\theta e^{\theta-1} + e^{\theta})C \geq 0.
\end{aligned}$$

By rearranging, we have

$$\begin{aligned}
C \leq \frac{1}{2(1-\alpha)\theta e^{\theta-1} + e^{\theta}}\Bigg(&(1-\theta)e^{(1-\alpha)\theta-1}+\\
&\frac{e^{(1-\alpha)\theta-1}(\alpha(\theta-2)+1) + e^{\theta-1}(2\alpha-1)}{\alpha^2}\Bigg).
\end{aligned}$$

Thus, in order to obtain the best approximation C, we need to maximize the right hand side of the inequality above over $\theta \in [0,1]$ and $\alpha \in [1/2, 1]$. Setting $\alpha = 1/2$ and $\theta = 0.18$ gives $C > 0.372$.

REFERENCES

[1] Niv Buchbinder, Moran Feldman, Joseph Naor, and Roy Schwartz. Submodular maximization with cardinality constraints. In *Proc. 25th SODA*, pages 1433–1452, 2014.

[2] Niv Buchbinder, Moran Feldman, Joseph (Seffi) Naor, and Roy Schwartz. A tight linear time (1/2)-approximation for unconstrained submodular maximization. In *Proc. 53nd FOCS*, 2012.

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

[4] Chandra Chekuri, T. S. Jayram, and Jan Vondrák. On multiplicative weight updates for concave and submodular function maximization. In *Proc. 6th ITCS*, pages 201–210, 2015.

[5] Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. Submodular function maximization via the multilinear relaxation and contention resolution schemes. *SIAM J. Comput.*, 43(6):1831–1879, 2014.

[6] Shaddin Dughmi, Tim Roughgarden, and Mukund Sundararajan. Revenue submodularity. *Theory of Computing*, 8(1):95–119, 2012.

[7] Moran Feldman, Joseph Naor, and Roy Schwartz. A unified continuous greedy algorithm for submodular maximization. In *Proc. 52nd FOCS*, pages 570–579, 2011.

[8] Shayan Oveis Gharan and Jan Vondrák. Submodular maximization by simulated annealing. In *Proc. 22nd SODA*, pages 1098–1117. SIAM, 2011.

[9] Ryan Gomes and Andreas Krause. Budgeted nonparametric learning from data streams. In *Proc. 27th ICML*, pages 391–398, 2010.

[10] Stefanie Jegelka and Jeff A. Bilmes. Submodularity beyond submodular energies: Coupling edges in graph cuts. In *The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011*, pages 1897–1904, 2011.

[11] David Kempe, Jon M. Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proc. 9th KDD*, pages 137–146, 2003.

[12] Andreas Krause, Ajit Paul Singh, and Carlos Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9:235–284, 2008.

[13] Jon Lee, Vahab S Mirrokni, Viswanath Nagarajan, and Maxim Sviridenko. Non-monotone submodular maximization under matroid and knapsack constraints. In *Proc. 41st STOC*, pages 323–332. ACM, 2009.

[14] Jon Lee, Maxim Sviridenko, and Jan Vondrák. Submodular maximization over multiple matroids via generalized exchange properties. *Mathematics of Operations Research*, 35(4):795–806, 2010.

[15] Hui Lin and Jeff A. Bilmes. Multi-document summarization via budgeted maximization of submodular functions. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 2-4, 2010, Los Angeles, California, USA*, pages 912–920, 2010.

[16] G L Nemhauser and L A Wolsey. Best algorithms for approximating the maximum of a submodular set function. *Mathematics of Operations Research*, 3(3):177–188, 1978.

[17] Maxim Sviridenko. A note on maximizing a submodular set function subject to a knapsack constraint. *Operations Research Letters*, 32(1):41–43, 2004.

[18] Jan Vondrák. Optimal approximation for the submodular welfare problem in the value oracle model. In *Proc. 40th STOC*, pages 67–74. ACM, 2008.

APPENDIX

In this section, we analyze Algorithm 0 given in Section III. We start by showing that the problem $\max_{\mathbf{u} \leq \mathbf{x} \leq \mathbf{v}} F(\mathbf{x})$ has an optimal solution OPT with the following property.

Lemma 12. *There is an optimal solution* OPT *to the problem* $\max_{\mathbf{u} \leq \mathbf{x} \leq \mathbf{v}} F(\mathbf{x})$ *such that for all* i, *either* $\text{OPT}_i = \mathbf{u}_i$ *or* $\text{OPT}_i = \mathbf{v}_i$.

Proof: Let $\text{OPT} = \arg\max_{\mathbf{u} \leq \mathbf{x} \leq \mathbf{v}} F(\mathbf{x})$ be an arbitrary optimal solution. Note that we can write each OPT_i as a convex combination of $\mathbf{u}_i$ and $\mathbf{v}_i$: $\text{OPT}_i = \gamma_i \mathbf{u}_i + (1 - \gamma_i)\mathbf{v}_i$, where $\gamma_i = (\text{OPT}_i - \mathbf{u}_i)/(\mathbf{v}_i - \mathbf{u}_i)$ if $\mathbf{u}_i \neq \mathbf{v}_i$ and $\gamma_i = 1$ otherwise. Let $\mathbf{o}_i$ be a random variable that is equal to $\mathbf{u}_i$ with probability γ_i and is equal to $\mathbf{v}_i$ with probability $1 - \gamma_i$. Let $\mathbf{o}$ be the vector whose coordinates are $\mathbf{o}_i$'s. By the definition of the multilinear extension, we have $F(\text{OPT}) = \mathbb{E}_{\mathbf{o}}[F(\mathbf{o})]$. Thus, there exists a realization $\mathbf{o} = \hat{\mathbf{o}}$ such that $F(\hat{\mathbf{o}}) \geq F(\text{OPT})$. Thus there is an optimal solution such that for all i, its ith coordinate is either $\mathbf{u}_i$ or $\mathbf{v}_i$. ∎

We will also use the following observation that follows from the definition of the multilinear extension.

Claim 13. *Let* $\mathbf{x} \in [0,1]^n$ *and* $\delta \in [-\mathbf{x}_i, 1 - \mathbf{x}_i]$. *We have*

$$F(\mathbf{x} + \delta \cdot \mathbf{1}_{\{i\}}) - F(\mathbf{x}) = \delta \cdot (F(\mathbf{x} \vee \mathbf{1}_{\{i\}}) - F(\mathbf{x} \wedge \mathbf{1}_{V \setminus \{i\}})).$$

Proof: Note that, for each $\mathbf{y} \in [0,1]^n$ and each $j \in [n]$,

$$F(\mathbf{y}) = \mathbf{y}_j \cdot F(\mathbf{y} \vee \mathbf{1}_{\{j\}}) + (1 - \mathbf{y}_j) \cdot F(\mathbf{y} \wedge \mathbf{1}_{V \setminus \{j\}}).$$

Therefore

$$F(\mathbf{x} + \delta \cdot \mathbf{1}_{\{i\}}) = (\mathbf{x}_i + \delta) \cdot F(\mathbf{x} \vee \mathbf{1}_{\{i\}}) + (1 - \mathbf{x}_i - \delta) \cdot F(\mathbf{x} \wedge \mathbf{1}_{V \setminus \{i\}}),$$
$$F(\mathbf{x}) = \mathbf{x}_i \cdot F(\mathbf{x} \vee \mathbf{1}_{\{i\}}) + (1 - \mathbf{x}_i) \cdot F(\mathbf{x} \wedge \mathbf{1}_{V \setminus \{i\}}),$$

and the claim follows. ∎

Proof of Lemma 5: Let $\text{OPT} = \arg\max_{\mathbf{u} \leq \mathbf{x} \leq \mathbf{v}} F(\mathbf{x})$. By Lemma 12, we may assume that for each i, either $\text{OPT}_i =$

$\mathbf{u}_i$ or $\mathrm{OPT}_i = \mathbf{v}_i$. Let $\mathrm{OPT}^{(i)} = (\mathrm{OPT} \vee \mathbf{u}^{(i)}) \wedge \mathbf{v}^{(i)}$. Note that $\mathbf{u}^{(i)} \leq \mathrm{OPT}^{(i)} \leq \mathbf{v}^{(i)}$ and therefore $\nabla F(\mathbf{u}^{(i)}) \geq \nabla F(\mathrm{OPT}^{(i)}) \geq \nabla F(\mathbf{v}^{(i)})$ by Fact 3.

We will show that

$$F(\mathrm{OPT}^{(i-1)}) - F(\mathrm{OPT}^{(i)}) \leq$$
$$\frac{1}{2}\left(F(\mathbf{u}^{(i)}) - F(\mathbf{u}^{(i-1)}) + F(\mathbf{v}^{(i)}) - F(\mathbf{v}^{(i-1)})\right). \quad (1)$$

Note that (1) immediately implies the lemma. We prove (1) in the following.

Suppose that $a_i' + b_i' \neq 0$. By Claim 13,

$$F(\mathbf{u}^{(i)}) - F(\mathbf{u}^{(i-1)})$$
$$= \frac{a_i'}{a_i' + b_i'}(\mathbf{v}_i - \mathbf{u}_i)\cdot$$
$$\left(F(\mathbf{u}^{(i-1)} \vee \mathbf{1}_{\{i\}}) - F(\mathbf{u}^{(i-1)} \wedge \mathbf{1}_{V \setminus \{i\}})\right)$$
$$= \frac{(a_i')^2(\mathbf{v}_i - \mathbf{u}_i)}{a_i' + b_i'}$$

and

$$F(\mathbf{v}^{(i)}) - F(\mathbf{v}^{(i-1)})$$
$$= \frac{b_i'}{a_i' + b_i'}(\mathbf{v}_i - \mathbf{u}_i)\cdot$$
$$\left(-F(\mathbf{v}^{(i-1)} \vee \mathbf{1}_{\{i\}}) + F(\mathbf{v}^{(i-1)} \wedge \mathbf{1}_{V \setminus \{i\}})\right)$$
$$= \frac{(b_i')^2(\mathbf{v}_i - \mathbf{u}_i)}{a_i' + b_i'}.$$

Now recall that we have either $\mathrm{OPT}_i = \mathbf{u}_i$ or $\mathrm{OPT}_i = \mathbf{v}_i$. If $\mathrm{OPT}_i = \mathbf{u}_i$, we have

$$F(\mathrm{OPT}^{(i-1)}) - F(\mathrm{OPT}^{(i)})$$
$$= \frac{a_i'}{a_i' + b_i'}(\mathbf{v}_i - \mathbf{u}_i)\cdot$$
$$\left(-F(\mathrm{OPT}^{(i-1)} \vee \mathbf{1}_{\{i\}}) + F(\mathrm{OPT}^{(i-1)} \wedge \mathbf{1}_{(V \setminus \{i\})})\right)$$
$$= -\frac{a_i'}{a_i' + b_i'}(\mathbf{v}_i - \mathbf{u}_i)\frac{\partial F}{\partial x_i}(\mathrm{OPT}^{(i-1)})$$
$$\leq -\frac{a_i'}{a_i' + b_i'}(\mathbf{v}_i - \mathbf{u}_i)\frac{\partial F}{\partial x_i}(\mathbf{v}^{(i-1)})$$
$$= \frac{a_i' b_i(\mathbf{v}_i - \mathbf{u}_i)}{a_i' + b_i'}$$
$$\leq \frac{a_i' b_i'(\mathbf{v}_i - \mathbf{u}_i)}{a_i' + b_i'}$$

On the first two lines, we have used Claim 13 and Fact 2. On the third line, we have used the fact that $\nabla F(\mathrm{OPT}^{(i-1)}) \geq \nabla F(\mathbf{v}^{(i-1)})$. On the fourth and fifth lines, we have used the definition of b_i and b_i'.

If $\mathrm{OPT}_i = \mathbf{v}_i$, we use an analogous argument.

$$F(\mathrm{OPT}^{(i-1)}) - F(\mathrm{OPT}^{(i)})$$

$$= \frac{b_i'}{a_i' + b_i'}(\mathbf{v}_i - \mathbf{u}_i)\cdot$$
$$(F(\mathrm{OPT}^{(i-1)} \vee \mathbf{1}_{\{i\}}) - F(\mathrm{OPT}^{(i-1)} \wedge \mathbf{1}_{V \setminus \{i\}}))$$
$$= \frac{b_i'}{a_i' + b_i'}(\mathbf{v}_i - \mathbf{u}_i)\frac{\partial F}{\partial x_i}(\mathrm{OPT}^{(i-1)})$$
$$\leq \frac{b_i'}{a_i' + b_i'}(\mathbf{v}_i - \mathbf{u}_i)\frac{\partial F}{\partial x_i}(\mathbf{u}^{(i-1)})$$
$$\leq \frac{a_i' b_i'(\mathbf{v}_i - \mathbf{u}_i)}{a_i' + b_i'}$$

Since $2a_i' b_i' \leq (a_i')^2 + (b_i')^2$, the inequality (1) follows.

Finally, suppose that $a_i' + b_i' = 0$. Notice that $a_i \geq -b_i$ by submodularity so this case can only happen if $a_i = b_i = 0$. In this case, we can set the ith coordinate of $\mathbf{u}$ and $\mathbf{v}$ to an arbitrary common value and by the same argument as above, we have $F(\mathbf{u}^{(i)}) = F(\mathbf{u}^{(i-1)}), F(\mathbf{v}^{(i)}) = F(\mathbf{v}^{(i-1)})$, and $F(\mathrm{OPT}^{(i)}) = F(\mathrm{OPT}^{(i-1)})$. Therefore (1) is trivially satisfied. $\qquad \square$

In Figure 3, we give a discretized version of the algorithm in Figure 2. In the remainder of this section, we show how to modify the analysis from Section V.

We discretize the time interval $[0, 1]$ into segments of size $\delta = n^{-4}$.

Analysis of the solution $\mathbf{y}^{(1)}$.: We modify the analysis from Subsection V-A as follows. (Note that the rest of the analysis remains unchanged.)

Consider an iteration θ of the outer for loop of the algorithm in Figure 3. In the remainder of this section, all of the vectors $\mathbf{x}^{(\cdot)}$, $\mathbf{y}^{(\cdot)}$, etc. refer to the vectors during iteration θ. All of the time steps are implicitly assumed to be the discrete time steps $\{0, \delta, 2\delta, 3\delta, \ldots, 1\}$.

First we prove bounds on $\|\mathbf{x}^{(t)}\|_\infty$ and $\|\mathbf{y}^{(t)}\|_\infty$ with a similar argument to Lemma 7. Consider a fixed $i \in [n]$. Since $\mathbf{v}_i^{(t)} \leq \alpha$ for all $t < \theta$, we have $1 - \mathbf{x}_i^{(t+\delta)} \geq (1 - \delta\alpha)(1 - \mathbf{x}_i^{(t)})$ for all $t < \theta$. Thus,

$$1 - \mathbf{x}_i^{(t)} \geq (1 - \delta\alpha)^{t/\delta}.$$

Similarly, for all $t \in [\theta, 1]$ and t a multiple of δ,

$$1 - \mathbf{y}_i^{(t)} \geq (1 - \delta\alpha)^{\theta/\delta}(1 - \delta)^{(t-\theta)/\delta}.$$

Next we prove a lower bound on $F(\mathbf{y}^{(1)})$ with a similar argument to Theorem 6.

Lemma 14. *Consider $\mathbf{u}, \mathbf{v}$ satisfying $|\mathbf{u}_i - \mathbf{v}_i| \leq \delta$. Then*

$$F(\mathbf{v}) - F(\mathbf{u}) \leq \delta n^2 M,$$

where $M = \max_{i \in [n]} f(\{i\})$.

Proof: First consider the case $\mathbf{u}, \mathbf{v}$ agree on all but coordinate i_0. Let $R(\mathbf{x})$ be the random set where each element i is independently included with probability $\mathbf{x}_i$. This

```
 1: Initialize $\mathbf{x}^* = 0$
 2: $\delta = n^{-4}$
 3: for $\theta \in \{0, \delta, 2\delta, 3\delta, \ldots, 1\}$ do
        (▷) Dampened Continuous Greedy stage
 4:     Initialize $\mathbf{x}^{(0)} = \mathbf{0}$
 5:     $t = 0$
 6:     while $t < \theta$ do
 7:         $\mathbf{v}^{(t)} = \mathrm{argmax}_{\mathbf{c} \in C, \|\mathbf{c}\|_\infty \leq \alpha} \langle \nabla F(\mathbf{x}^{(t)}) \circ (\mathbf{1} - \mathbf{x}^{(t)}), \mathbf{c} \rangle$
 8:         $\mathbf{x}^{(t+\delta)} = \mathbf{x}^{(t)} + \delta \mathbf{v}^{(t)} \circ (\mathbf{1} - \mathbf{x}^{(t)})$
 9:         $t = t + \delta$
10:     end while

        (▷) Standard Continuous Greedy stage
11:     Initialize $\mathbf{y}^{(\theta)} = \mathbf{x}^{(\theta)}$
12:     while $t < 1$ do
13:         $\mathbf{v}^{(t)} = \mathrm{argmax}_{\mathbf{c} \in C} \langle \nabla F(\mathbf{y}^{(t)}) \circ (\mathbf{1} - \mathbf{y}^{(t)}), \mathbf{c} \rangle$
14:         $\mathbf{y}^{(t+\delta)} = \mathbf{y}^{(t)} + \delta \mathbf{v}^{(t)} \circ (\mathbf{1} - \mathbf{y}^{(t)})$
15:         $t = t + \delta$
16:     end while

        (▷) Double Greedy stage
17:     $\mathbf{p} = \mathrm{argmax}_{\mathbf{c} \in C} \langle \nabla F(\mathbf{x}^{(\theta)}) \circ (\mathbf{1} - \mathbf{x}^{(\theta)}), \mathbf{c} \rangle$
18:     Find $\mathbf{z}$ approximating $\mathrm{argmax}_{\mathbf{c} \leq \mathbf{p}} F(\mathbf{c})$ using
        the Double Greedy algorithm (Figure 1)

        (▷) Update the best solution
19:     if $F(\mathbf{y}^{(1)}) > F(\mathbf{x}^*)$ then $\mathbf{x}^* = \mathbf{y}^{(1)}$
20:     end if
21:     if $F(\mathbf{z}) > F(\mathbf{x}^*)$ then $\mathbf{x}^* = \mathbf{z}$
22:     end if
23: end for
24: return $\mathbf{x}^*$
```

Figure 3. Discretized algorithm for a general constraint

process can be thought of as picking a random $r_i \in [0, 1]$ and including i if $r_i \leq \mathbf{x}_i$. One can generate (coupled) $R(\mathbf{u})$ and $R(\mathbf{v})$ by sharing the same r_i's. Notice that $R(\mathbf{u})$ and $R(\mathbf{v})$ agree on all coordinates other than i_0 and they disagree on coordinate i_0 with probability at most δ. Thus we have

$$\mathbb{E}[f(R(\mathbf{u})) - f(R(\mathbf{v}))] \leq \delta \mathbb{E}[f(R(\mathbf{u})) \mid R(\mathbf{u}) \neq R(\mathbf{v})]$$
$$\leq \delta \max_{S \subseteq V} f(S) \leq \delta n M.$$

Next, let $\mathbf{u}^i$ be the vector whose first i coordinates agree with $\mathbf{u}$ and the last $n - i$ coordinates agree with $\mathbf{v}$. We have

$$F(\mathbf{v}) - F(\mathbf{u}) = \sum_{i=0}^{n-1} (F(\mathbf{u}^i) - F(\mathbf{u}^{i+1})) \leq \delta n^2 M,$$

where the inequality comes from the above special case. ∎

Corollary 15. *Consider* $\mathbf{u}$ *and* $\mathbf{v}$ *satisfying* $\mathbf{u} \leq \mathbf{v}$ *and* $\mathbf{v}_i \leq \mathbf{u}_i + \delta$. *Then*

$$\frac{\partial F}{\partial x_i}(\mathbf{u}) - \frac{\partial F}{\partial x_i}(\mathbf{v}) \leq 2\delta n^2 M.$$

Proof: By Claim 2,

$$\frac{\partial F}{\partial x_i}(\mathbf{u}) - \frac{\partial F}{\partial x_i}(\mathbf{v}) = F(\mathbf{u} \vee \mathbf{1}_{\{i\}}) - F(\mathbf{v} \vee \mathbf{1}_{\{i\}}) -$$
$$F(\mathbf{u} \wedge \mathbf{1}_{V \setminus \{i\}}) + F(\mathbf{v} \wedge \mathbf{1}_{V \setminus \{i\}}) \leq 2\delta n^2 M.$$

∎

For any given $t \in [\theta, 1 - \delta]$, we have

$$F(\mathbf{y}^{(t+\delta)}) - F(\mathbf{y}^{(t)})$$
$$= \int_0^1 \langle \nabla F((1-z)\mathbf{y}^{(t)} + z\mathbf{y}^{(t+\delta)}), \mathbf{y}^{(t+\delta)} - \mathbf{y}^{(t)} \rangle dz$$
$$= \langle \nabla F(\mathbf{y}^{(t)}), \mathbf{y}^{(t+\delta)} - \mathbf{y}^{(t)} \rangle +$$
$$\int_0^1 \Big\langle \nabla F((1-z)\mathbf{y}^{(t)} + z\mathbf{y}^{(t+\delta)}) - \nabla F(\mathbf{y}^{(t)}),$$
$$\mathbf{y}^{(t+\delta)} - \mathbf{y}^{(t)} \Big\rangle dz$$
$$\geq \langle \nabla F(\mathbf{y}^{(t)}), \mathbf{y}^{(t+\delta)} - \mathbf{y}^{(t)} \rangle -$$
$$\int_0^1 \| \nabla F((1-z)\mathbf{y}^{(t)} + z\mathbf{y}^{(t+\delta)}) - \nabla F(\mathbf{y}^{(t)}) \|_\infty \cdot$$
$$\| \mathbf{y}^{(t+\delta)} - \mathbf{y}^{(t)} \|_1 dz$$
$$\geq \langle \nabla F(\mathbf{y}^{(t)}), \mathbf{y}^{(t+\delta)} - \mathbf{y}^{(t)} \rangle - \delta^2 n^3 M$$
$$= \langle \nabla F(\mathbf{y}^{(t)}), \delta \mathbf{v}^{(t)} \circ (\mathbf{1} - \mathbf{y}^{(t)}) \rangle - \delta^2 n^3 M$$
$$= \langle \nabla F(\mathbf{y}^{(t)}) \circ (\mathbf{1} - \mathbf{y}^{(t)}), \delta \mathbf{v}^{(t)} \rangle - \delta^2 n^3 M$$
$$\geq \langle \nabla F(\mathbf{y}^{(t)}) \circ (\mathbf{1} - \mathbf{y}^{(t)}), \delta \mathrm{OPT} \rangle - \delta^2 n^3 M$$
$$\geq \delta(F(\mathbf{y}^{(t)} \vee \mathrm{OPT}) - F(\mathbf{y}^{(t)})) - \delta^2 n^3 M$$
$$\geq \delta(1 - \delta\alpha)^{\theta/\delta}(1 - \delta)^{(t-\theta)/\delta} \cdot F(\mathrm{OPT}) - \delta \cdot F(\mathbf{y}^{(t)}) -$$
$$\delta^2 n^3 M \quad \text{(By Lemma 10)}$$

Therefore we have the following recurrence for $F(\mathbf{y}^{(t)})$:

$$F(\mathbf{y}^{(t+\delta)}) \geq (1 - \delta)F(\mathbf{y}^{(t)}) +$$
$$\delta(1 - \delta\alpha)^{\theta/\delta}(1 - \delta)^{(t-\theta)/\delta} F(\mathrm{OPT}) - \delta^2 n^3 M$$

By expanding the recurrence, we obtain

$$F(\mathbf{y}^{(t)}) \geq (1 - \delta)^{(t-\theta)/\delta} \Big((t - \theta)(1 - \delta\alpha)^{\theta/\delta} \cdot F(\mathrm{OPT}) +$$
$$F(\mathbf{x}^{(\theta)}) \Big) - t\delta n^3 M.$$

By substituting $t = 1$ and $\delta = n^{-4}$ and by using the inequalities $e^{-x/(1-x)} \leq 1 - x \leq e^{-x}$ for all $0 < x < 1$, we obtain

$$F(\mathbf{y}^{(1)}) \geq e^{-1+\theta} \Big((1 - \theta)e^{-\alpha\theta} F(\mathrm{OPT}) + F(\mathbf{x}^{(\theta)}) \Big) -$$
$$O(n^{-1}) F(\mathrm{OPT})$$

$$=\frac{1}{e}\left((1-\theta)e^{(1-\alpha)\theta}F(\text{OPT})+e^{\theta}F(\mathbf{x}^{(\theta)})\right)-$$
$$O(n^{-1})F(\text{OPT}).$$

Combining the two solutions $\mathbf{y}^{(1)}$ and $\mathbf{z}$.: We extend the argument in Subsection V-C as follows. As we have shown above, in every iteration θ that is a multiple of δ, we obtain two solutions $\mathbf{y}^{(1)}$ and $\mathbf{z}$ satisfying:

$$F(\mathbf{y}^{(1)})\geq\frac{1}{e}\left((1-\theta)e^{(1-\alpha)\theta}F(\text{OPT})+e^{\theta}F(\mathbf{x}^{(\theta)})\right)-$$
$$O(n^{-1})F(\text{OPT})$$
$$F(\mathbf{z})\geq\frac{1}{2(1-\alpha)}\left(e^{-\alpha\theta}F(\text{OPT})-F(\mathbf{x}^{(\theta)})-\right.$$
$$\left.\langle\nabla F(\mathbf{x}^{(\theta)})\circ(\mathbf{1}-\mathbf{x}^{(\theta)}),\mathbf{v}^{(\theta)}\rangle\right)$$

As before, we show that there is an iteration θ for which $\max\{F(\mathbf{y}^{(1)}),F(\mathbf{z})\}>C\cdot F(\text{OPT})$, where C is a constant that we will set later. We will proceed by contradiction and assume that $\max\{F(\mathbf{y}^{(1)}),F(\mathbf{z})\}\leq C\cdot F(\text{OPT})$ for all θ. We take a convex combination of the two inequalities and eliminate $F(\mathbf{x}^{(\theta)})$. Thus we get that, for all θ,

$$\frac{1}{\frac{1}{2(1-\alpha)}+e^{\theta-1}}\left(\frac{(2-\theta)e^{(1-\alpha)\theta-1}}{2(1-\alpha)}F(\text{OPT})-\right.$$
$$\left.\frac{e^{\theta-1}}{2(1-\alpha)}\langle\nabla F(\mathbf{x}^{(\theta)})\circ(\mathbf{1}-\mathbf{x}^{(\theta)}),\mathbf{v}^{(\theta)}\rangle\right)\leq C'\cdot F(\text{OPT})$$

where $C'=C+O(n^{-1})$. Thus, for all $\theta\in[0,1]$,

$$\langle\nabla F(\mathbf{x}^{(\theta)})\circ(\mathbf{1}-\mathbf{x}^{(\theta)}),\mathbf{v}^{(\theta)}\rangle\geq$$
$$\left((2-\theta)e^{-\alpha\theta}-(e^{1-\theta}+2(1-\alpha))C'\right)F(\text{OPT})$$

For any given $t\in[0,\theta)$, we have

$$F(\mathbf{x}^{(t+\delta)})-F(\mathbf{x}^{(t)})$$
$$=\int_{0}^{1}\langle\nabla F((1-z)\mathbf{x}^{(t)}+z\mathbf{x}^{(t+\delta)}),\mathbf{x}^{(t+\delta)}-\mathbf{x}^{(t)}\rangle dz$$
$$=\langle\nabla F(\mathbf{x}^{(t)}),\mathbf{x}^{(t+\delta)}-\mathbf{x}^{(t)}\rangle+$$
$$\int_{0}^{1}\left\langle\nabla F((1-z)\mathbf{x}^{(t)}+z\mathbf{x}^{(t+\delta)})-\nabla F(\mathbf{x}^{(t)}),\right.$$
$$\left.\mathbf{x}^{(t+\delta)}-\mathbf{x}^{(t)}\right\rangle dz$$
$$\geq\langle\nabla F(\mathbf{x}^{(t)}),\mathbf{x}^{(t+\delta)}-\mathbf{x}^{(t)}\rangle-$$
$$\int_{0}^{1}\|\nabla F((1-z)\mathbf{x}^{(t)}+z\mathbf{x}^{(t+\delta)})-\nabla F(\mathbf{x}^{(t)})\|_{\infty}\cdot$$
$$\|\mathbf{x}^{(t+\delta)}-\mathbf{x}^{(t)}\|_{1}dz$$
$$\geq\langle\nabla F(\mathbf{x}^{(t)}),\mathbf{x}^{(t+\delta)}-\mathbf{x}^{(t)}\rangle-\delta^{2}n^{3}M$$
$$=\langle\nabla F(\mathbf{x}^{(t)}),\delta\mathbf{v}^{(t)}\circ(\mathbf{1}-\mathbf{x}^{(t)})\rangle-\delta^{2}n^{3}M$$
$$=\delta\langle\nabla F(\mathbf{x}^{(t)})\circ(\mathbf{1}-\mathbf{x}^{(t)}),\delta\mathbf{v}^{(t)}\rangle-\delta^{2}n^{3}M$$
$$\geq\delta\left((2-t)e^{-\alpha t}-(e^{1-t}+2(1-\alpha))C'\right)F(\text{OPT})-$$

$$\delta^{2}n^{3}M$$

By expanding the recurrence, we obtain

$$F(\mathbf{x}^{(t+\delta)})\geq\sum_{i=0}^{t/\delta}\left(\delta\left((2-t+i\delta)e^{-\alpha(t-i\delta)}-(e^{1-t+i\delta}+\right.\right.$$
$$\left.\left.2(1-\alpha))C'\right)F(\text{OPT})-\delta^{2}n^{3}M\right) \qquad (2)$$

By computation, we have

$$\sum_{i=0}^{t/\delta}\delta(2-t+i\delta)e^{-\alpha(t-i\delta)}$$
$$=\delta e^{-\alpha t}\left((2-t-\delta)\frac{e^{\alpha\delta(t/\delta+1)}-1}{e^{\alpha\delta}-1}+\right.$$
$$\left.\delta\frac{(t/\delta+1)e^{\alpha\delta(t/\delta+2)}-(t/\delta+2)e^{\alpha\delta(t/\delta+1)}+1}{(e^{\alpha\delta}-1)^{2}}\right)$$
$$\geq e^{-\alpha t}\left(\frac{(2-t)(e^{\alpha t}-1)}{\alpha}+\frac{t\alpha e^{\alpha t}-e^{\alpha t}+1}{\alpha^{2}}\right)-O(\delta)$$
$$=\frac{(t-2)\alpha e^{-\alpha t}+2\alpha-1+e^{-\alpha t}}{\alpha^{2}}-O(\delta)$$

We also have

$$\sum_{i=0}^{t/\delta}\delta(e^{1-t+i\delta}+2(1-\alpha))C'$$
$$=\left(\delta e^{1-t}\frac{e^{t+\delta}-1}{e^{\delta}-1}+2(1-\alpha)(t+\delta)\right)C'$$
$$\leq\left(e-e^{1-t}+2(1-\alpha)(t+\delta)\right)C'+O(\delta)$$

Substituting into Equation 2, we have

$$F(\mathbf{x}^{(t+\delta)})\geq\left(\frac{(t-2)\alpha e^{-\alpha t}+2\alpha-1+e^{-\alpha t}}{\alpha^{2}}-(e-\right.$$
$$\left.e^{1-t}+2(1-\alpha)t)C'-O(\delta n^{3})\right)F(\text{OPT})$$

Notice that the RHS is exactly the same as in the continuous argument except for an additive error $O(\delta n^{3})F(\text{OPT})$. By carrying this error through, we can finish the proof exactly the same as before.