Current Status and Planned Next Steps

1. Current Status of the Formalisation

The development is currently split into two theory files:

Greedy_Submodular_Construct

- Established the locale `Greedy_Setup`, fixing a finite ground set V, a budget k, and a non-negative, monotone, submodular set function f with $f(\emptyset) = 0$.
- Defined the marginal gain $\text{gain}(S, e)$, the maximiser argmax_gain, and the greedy sequence

$$\text{greedy\_set} : \mathbb{N} \Rightarrow \text{“sets of items”}.$$

- Proved fundamental structural properties:
  - $\text{greedy\_set}(i) \subseteq V$ and every greedy set is finite.
  - Monotonicity: $\text{greedy\_set}(i) \subseteq \text{greedy\_set}(j)$ for $i \leq j$.
  - Cardinality bounds: $\text{card}(\text{greedy\_set}(i)) \leq i$ and $\text{card}(\text{greedy\_set}(i)) \leq \min(i, \text{card}(V))$.
  - A precise state-transition lemma describing the evolution of $\text{greedy\_set}(i+1)$ when the remainder $V - \text{greedy\_set}(i)$ is non-empty.
- Defined a list view of the greedy construction (`greedy_sequence`) and established indexing lemmas.

Greedy_Submodular_Approx

- Proved the analytic inequality for all $k \geq 1$:

$$(1 - 1/k)^k \leq e^{-1} \quad \text{and} \quad 1 - (1 - 1/k)^k \geq 1 - 1/e.$$

- Within the locale:
  - Established non-negativity of marginal gains and two basic non-emptiness lemmas.
  - Proved diminishing returns and the submodular telescoping inequality.
  - Proved the averaging lemma ensuring the existence of an element $e \in V \setminus S$ with
$$\text{gain}(S, e) \geq \frac{f(\text{Opt}) - f(S)}{k}.$$
  - Collected feasible sets $\mathcal{F}_k = \{S \subseteq V : |S| \leq k\}$, defined a canonical maximiser OPT_set, and set $\text{OPT}_k = f(\text{OPT\_set})$.

- Introduced the gap sequence:

$$\mathrm{gap}(i) = \mathrm{OPT}_k - f(\mathrm{greedy\_set}(i)),$$

  and showed the linear recurrence and geometric decay:

$$\mathrm{gap}(i) \leq (1 - 1/k)^i \cdot \mathrm{OPT}_k.$$

- Derived the Nemhauser–Wolsey guarantee:

$$f(\mathrm{greedy\_set}(k)) \geq \left(1 - (1 - 1/k)^k\right) \cdot \mathrm{OPT}_k \geq (1 - 1/e)\, \mathrm{OPT}_k.$$

- Defined an approximation-ratio function and proved it is always at least $1 - 1/e$.

This completes the full formal development of the classical greedy approximation bound.

2. Planned Next Steps

At this stage, it seems natural to shift the focus toward applications, reusable instantiations, and small executable experiments. I believe the following directions could form a reasonable next phase, and, if successful, would help turn the development into a reusable and practically meaningful component.

A. Coverage Function Instantiation

- I plan to introduce a locale `Coverage_Setup` with:
  - a finite universe U,
  - a ground set V,
  - a mapping $g : V \Rightarrow \mathcal{P}(U)$,
  - a cardinality budget k.
- Then I'll define the coverage function

  $$f_{\mathrm{cov}}(S) = \left| \bigcup_{x \in S} g(x) \right|,$$

  which, in principle, should serve as a canonical example of a monotone submodular function.
- I expect that it should be possible to show that f_{cov} satisfies the assumptions of `Greedy_Setup`, namely:
  - non-negativity and $f_{\mathrm{cov}}(\emptyset) = 0$,
  - monotonicity under inclusion,
  - submodularity (likely via the standard union–intersection argument).
- If these proofs go through as anticipated, I should be able to use `interpret` to instantiate the greedy theory with f_{cov}.
- From there, it seems reasonable to attempt deriving a coverage-specialised approximation theorem:

  $$f_{\mathrm{cov}}(\mathrm{greedy\_set}(k)) \geq (1 - 1/e)\,\mathrm{OPT}_k.$$

- I would also want to add a brief explanatory text section summarising the set-cover model and why this instantiation is meaningful.

B. Baseline: Exhaustive Search

- I think it could be useful to define an exhaustive maximiser such as

$$\text{enum\_opt}(V) = \arg \max_{S \subseteq V,\ |S| \leq k} f(S),$$

especially for small finite instances.

- My plan is to prove that this construction indeed yields a feasible maximiser with value OPT_k, this should be relatively straightforward since the feasible family is finite.

- I believe it might also be helpful to provide a short informal complexity comparison:

$$\text{greedy: } O(k \cdot |V|), \qquad \text{exhaustive: } O(2^{|V|}),$$

to highlight the difference in scalability.

- (Optional) If time permits, I may also include a tiny concrete example illustrating the gap between greedy and the exact optimum.

C. Non-Submodular Counterexample

- I plan to define a simple non-submodular function, for example

$$f_{\text{quad}}(S) = (\text{card}(S))^2,$$

which should already violate submodularity on small domains.

- I intend to give an explicit violation of submodularity to make the failure transparent.

- It might be possible to construct a small instance where greedy performs strictly worse than the optimal value by more than the $1 - 1/e$ factor.

- If that works out, I could then summarise the result in a lemma indicating that the submodularity assumption is genuinely necessary for the approximation theorem.

D. Executable Code and Small Experiments

- I plan to experiment with Isabelle's `export_code` to generate executable code for `greedy_set`, and possibly for the coverage instance as well.

- My idea is to evaluate greedy on a few small example instances and record the outputs, mainly to illustrate how the verified construction behaves computationally.

- I plan to summarise these observations in a short text/comment section, which could also be useful for future users of the development.

E. Documentation and Polishing

- I think it would be worthwhile to gradually refine the structure and explanatory text within the theories so that the development becomes easier to follow, especially for potential future users.
- I would also want to add a clearer high-level roadmap summarising the logical flow of the formalisation:

$$\begin{aligned} &\text{construction} \\ \to\ &\text{submodular calculus} \\ \to\ &\text{averaging lemma} \\ \to\ &\text{gap recurrence} \\ \to\ &\text{approximation theorem} \\ \to\ &\text{applications.} \end{aligned}$$

- I feel that normalising naming conventions and cleaning up unused auxiliary lemmas could make the development more consistent, which might be helpful if I later adapt the theories for a workshop submission or extend them into a more polished journal version.
- Finally, I hope to prepare a concise external summary (outside Isabelle) that could serve either as a short report or the seed of a future paper draft, depending on how the project evolves.

Longer-term outlook. In the longer term, I hope that these components could potentially grow into a small reusable Isabelle library for submodular optimisation, or perhaps form the basis of a short workshop submission that could later be refined into a journal-style version.