Optimal Search under Constraints

Martine Ceberio, Olga Kosheleva,
and Vladik Kreinovich

University of Texas at El Paso
El Paso, TX 79968, USA
mceberio@utep.edu, olgak@utep.edu
vladik@utep.edu

1. Need to Select Optimal Dose of a Medicine

- This research started with a simple observation about how medical doctors decide on the dosage.
- For many chronic health conditions like high cholesterol, high blood pressure, etc.:
  - there are medicines
  - that bring the corresponding numbers back to normal.
- An important question is how to select the correct dosage.
- On the one hand, if the dosage is too small, the medicine will not have the full desired effect.

2. Need to Select Optimal Dose (cont-d)

- On the other hand, we do not want the dosage to be higher than needed:
  - every medicine has negative side effects, side effects that increase with the increase in dosage, and
  - we want to keep these side effects as small as possible.
- In most such cases, there are general recommendations that:
  - provide a range of possible doses
  - depending on the patient's age, weight, etc.
- However, a specific dosage within this range:
  - has to be selected individually,
  - based on how this patient's organism reacts to this medicine.

3. How the First Doctor Selected the Dose

- It so happened that two people having similar conditions ended up with the same daily dosage of 137 units.
- However interestingly, their doctors followed a different path to this value.
- For the first patient, the doctor seems to have followed the usual bisection algorithm.
- This doctor started with the dose of 200 – and it worked.
- So, the doctor tried 100 – it did not work.
- The doctor tried 150 – it worked.
- The doctor tried 125 – it did not work.
- So, the doctor tried 137 – and it worked.

4. The First Doctor Selected the Dose (cont-d)

- The doctor could have probably continued further, but:
  - the pharmacy already had trouble with maintaining the exact dose of 137,
  - so this became the final arrangement.
- This procedure indeed follows the usual bisection (= binary search) algorithm.
- This algorithm is usually described as a way:
  - to solve the equation $f(x) = 0$
  - when we have an interval $[a, b]$ for which

$$f(a) < 0 < f(b).$$

5. The First Doctor Selected the Dose (cont-d)

- In our problem, $f(a)$ is the difference between the effect of the dose a and the desired effect:
  - if the dose is not sufficient, this difference is negative, and
  - if the dose is sufficient, this difference is non-negative (positive or 0).
- In the bisection algorithm, at each iteration, we have a range $[\underline{x}, \overline{x}]$ for which $f(\underline{x}) < 0$ and $f(\overline{x}) > 0$.
- In the beginning, we have $[\underline{x}, \overline{x}] = [a, b]$.
- At each iteration, we take a midpoint $m = \dfrac{\underline{x} + \overline{x}}{2}$ and compute $f(m)$.

6. The First Doctor Selected the Dose (cont-d)

- Depending on the sign of $f(m)$, we make the following changes:
  - if $f(m) < 0$, we replace $\underline{x}$ with m and thus, get a new interval $[m, \overline{x}]$;
  - if $f(m) > 0$, we replace $\overline{x}$ with m and thus, get a new interval $[\underline{x}, m]$.
- In both cases, we decrease the width of the interval $[\underline{x}, \overline{x}]$ by half.
- We stop when this width becomes smaller than some given value $\varepsilon > 0$.
- The value ε represents the accuracy with which we want to find the solution.
- In the above example, based on the 1st experiment, we know that the ideal dose is in the interval $[0, 200]$.

7. The First Doctor Selected the Dose (cont-d)

- So, we try $m = 100$ and:
  - after finding that $f(m) < 0$ (i.e., that the dose $m = 100$ is not sufficient),
  - we come up with the narrower interval $[100, 200]$.
- Then, we try the new midpoint $m = 150$, and:
  - based on the testing result,
  - we come up with the narrower interval $[100, 150]$.
- Then, we try the new midpoint $m = 125$, and:
  - based on the testing result,
  - we come up with the narrower interval $[125, 150]$.
- In the last step, we try the new midpoint $m = 137$.

8. The First Doctor Selected the Dose (cont-d)

- Strictly speaking, it should be 137.5.
- However, as we have mentioned, the pharmacy cannot provide such an accuracy.
- Now we know that the desired value is within the narrower interval $[125, 137]$.
- Out of all possible values from the interval $[125, 137]$:
  - the only value about which we know that this value is sufficient
  - is the value 137.
- So this value has been prescribed to the first patient.

9. The Second Doctor Selected the Same Dose Differently

- Interestingly, for the second patient, the process was completely different.
- The doctor started with 25 units.
- Then – since this dose was not sufficient – the dose was increased to 50 units.
- Then the dose was increased to 75, 100, 125 units, and, finally, to 150 units.
- The 150 units dose turned out to be sufficient.
- So the doctor knew that the optimal dose is between 125 and 150.
- Thus, this doctor tried 137, and it worked.

10. The Second Doctor Selected the Same Dose Differently (cont-d)

- Interestingly:
  - in contrast to the first doctor,
  - this doctor could not convince the pharmacy to produce a 137 units dose.
- So this doctor's prescription of this dose consists of taking 125 units and 150 units in turn.

11. Why the Difference?

- Why did the two doctors use different procedures?
- Clearly, the second doctor needed more steps – and longer time – to come up with the same optimal dose:
  - this doctor used 7 steps:

    $$(25, 50, 75, 100, 125, 150, 137)$$

  - instead of only 5 steps used by the first doctor:

    $$(200, 100, 150, 125, 137).$$

- Why did this doctor not use a faster bisection procedure?
- At first glance, it may seem that the second doctor was not familiar with bisection.

12. Why the Difference (cont-d)

- However, clearly this doctor *was* familiar with it; e.g.:
  - after realizing that the optimal dose is within the interval $[125, 150]$,
  - he/she checked the midpoint dose of 137.
- The real explanation of:
  - why the second doctor did not use the faster procedure
  - is that the second doctor was more cautious about possible side effects.
- Probably, in this doctor's opinion, the second patient was vulnerable to possible side effects.

13. Why the Difference (cont-d)

- Thus, this doctor decided:
  - not to increase the dose too much beyond the optimal value,
  - so as to minimize possible side effects.
- On the other hand, the first doctor:
  - based on the overall health of the first patient,
  - was less worried about possible side effects.

14. Natural General Question

- A natural next question is:
  - under such restriction on possible tested values x,
  - what is the optimal way to find the desired solution,
  - i.e., to be more precise, the desired ε-approximation to the solution?
- It is known that:
  - if we do not have any constraints,
  - then bisection is the optimal way to find the solution to the equation $f(x) = 0$.
- So, the question is – how to optimally modify bisection under such constraints?

15. Towards Formulating the Problem in Precise Terms

- The larger the dose of the medicine, the larger the effect.
- There is a certain threshold x_0 after which the medicine has the full desired curing effect.
- Every time we test a certain does x of the medicine of a patient:
  - we either get the full desired effect, which would mean that $x_0 \leq x$,
  - or we do not yet get the full desired effect, which means that $x < x_0$.
- We want to find the curing dose as soon as possible, i.e., after as few tests as possible.

16. Towards Precise Formulation (cont-d)

- If the only objective was to cure the disease, then we could use any dose larger than or equal to x_0.
- However, the larger the dose, the larger the undesired side effects.
- So, we would like to prescribe a value which is as close to x_0 as possible.
- Of course, in real life, we can only maintain the dose with some accuracy $\varepsilon > 0$.
- So, we want to prescribe a value x_r which is ε-close to x_0, i.e., for which $x_0 \leq x_r \leq x_0 + \varepsilon$.

17. Towards Precise Formulation (cont-d)

- The only way to find the optimal dose x_r is to test different doses on a given patient.
- If, during this testing, we assign too large a dose, we may seriously harm the patient.
- So, it is desirable not to exceed x_0 too much when testing.
- Let us denote the largest allowed excess by δ.
- This means that we can only test values $x \le x_0 + \delta$.
- Now, we can formulate the problem in precise terms.

18. Precise Formulation of the Problem

- By *search under constraints*, we mean the following problem:
- *Given:* rational numbers $\varepsilon > 0$ and $\delta > 0$, and an algorithm c that, for some fixed (unknown) $x_0 > 0$,
  – given a rational number $x \in [0, x_0 + \delta]$,
  – checks whether $x < x_0$ or $x \geq x_0$.
- This algorithm will be called a *checking* algorithm.
- *Find:* a real number x_r for which $x_0 \leq x_r \leq x_0 + \varepsilon$.

19. Precise Formulation of the Problem (cont-d)

- We want to find the fastest possible algorithm for solving this problem.
- To gauge the speed of this algorithm, we will count the number of calls to the checking algorithm c.
- For every algorithm A , let us denote the number of calls to c by $N_{A,\varepsilon,\delta}(x_0)$.
- We say that the algorithm A_0 is *optimal* if every other algorithm A: $N_{A_0,\varepsilon,\delta}(x_0) \le N_{A,\varepsilon,\delta}(x_0) + \text{const}$.

20. Optimal Algorithm

Proposition. *The following algorithm $\mathcal{A}$ is optimal:*

- *First, we apply the algorithm c to values δ, 2δ, $\ldots$, until we find a value i for which $i \cdot \delta < x_0 \le (i+1) \cdot \delta$.*
- *Then, we apply bisection process to the interval $[i \cdot \delta, (i+1) \cdot \delta]$ to find x_r.*
- *In this process, at each moment of time, we have an interval $[\underline{x}, \overline{x}]$ for which $\underline{x} < x_0 \le \overline{x}$.*
- *We start with $[\underline{x}, \overline{x}] = [i \cdot \delta, (i+1) \cdot \delta]$.*
- *At each iteration step, we apply the checking algorithm c to the midpoint $m = \dfrac{\underline{x} + \overline{x}}{2}$.*

21. Optimal Algorithm (cont-d)

- *If it turns out that* $m < x_0$, *we replace* $[\underline{x}, \overline{x}]$ *with* $[m, \overline{x}]$.
- *If it turns out that* $x_0 \le m$, *we replace* $[\underline{x}, \overline{x}]$ *with*
$$[\underline{x}, m].$$
- *In both cases, we decrease the width of the interval by 2.*
- *We stop when this width becomes smaller than or equal to* ε, *i.e., when* $\overline{x} - \underline{x} \le \varepsilon$.
- *Then, we take* $\overline{x}$ *as the desired output* x_r.

22. Proof

- It is easy to prove that the algorithm $\mathcal{A}$ indeed solves the search under constraints problem.
- Indeed, increasing the previously tested value $x \le x_0$ is legitimate: since then $x + \delta \le x_0 + \delta$.
- By this increase, for each x_0, we will eventually find the value i for which $x_0 \le (i+1)\cdot\delta$ – namely, $i = \left\lceil \dfrac{x_0}{\delta} \right\rceil - 1$.
- Then, by induction, we can prove that on each step of the bisection process, we indeed have $\underline{x} < x_0 \le \overline{x}$.
- And if $\underline{x} < x_0 \le \overline{x}$ and $\overline{x} - \underline{x} \le \varepsilon$, then indeed $x_0 \le x_r = \overline{x} \le \underline{x} + \delta < x_0 + \varepsilon$.
- Optimality if also easy to prove.
- Indeed, the algorithm $\mathcal{A}$ takes $\dfrac{x_0}{\delta}$ + const steps.

23. Proof (cont-d)

- Here the constant – approximately equal to $\log_2\left(\dfrac{\delta}{\varepsilon}\right)$ – covers the bisection part.
- Let us show that other algorithms A cannot use fewer steps.
- Indeed, if v is the largest value for which we have already checked that $v < x_0$, then:
  - at the next test,
  - we cannot use the value $x > v + \delta$.
- Indeed, in this case, we have $v < x - \delta$.
- So for any x_0 from the interval $(v, x - \delta)$, we have $v < x_0 < x - \delta$ and thus, $x > x_0 + \delta$.
- Hence, for this x_0, the checking algorithm c is not applicable.

24. Proof (cont-d)

- Thus, at each step:
  - we cannot increase the tested value x
  - by more than δ in comparison with the previously tested value.
- So, to get to a value $x \geq x_0$ – which is our goal – we need to make at least $\dfrac{x_0}{\delta}$ calls to c.
- The proposition is proven.

25. Comment

- This is exactly what the both doctors did.
- The difference is that:
  - the first doctor used $\delta = 200$, while
  - the second doctor used a much smaller value

$$\delta = 25.$$

26. Acknowledgments

This work was supported in part by the National Science Foundation grants:

- 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science),
- HRD-1242122 (Cyber-ShARE Center of Excellence).