Solution Methods – Numerical Algorithms

Evelien van der Hurk

DTU Managment Engineering

Class Exercises From Last Time

Class Exercise 2

$$\begin{aligned} \text{minimize:} \quad & -2x_1 + x_2 \\ \text{subject to:} \quad & x_1 + x_2 = 3 \\ & (x_1, x_2) \in X \end{aligned}$$

1. Suppose X={(0,0),(0,4),(4,4),(4,0),(1,2),(2,1)}
2. Formulate the Lagrangian Dual Problem
3. Plot the Lagrangian Dual Problem
4. Find the optimal solution to the primal and dual problems
5. Check whether the objective functions are equal
6. Explain your observation in 5

Solution (1)

- Let λ be the lagrange multiplier on the equality constraint
- Lagrangian is then:

$$L(\boldsymbol{x}, \lambda) = -2x_1 + x_2 + \lambda(3 - x_1 - x_2)$$

- The Lagrangian Dual Function is then:

$$\theta(\lambda) = \min_{\boldsymbol{x} \in X}\{-2x_1 + x_2 + \lambda(3 - x_1 - x_2)\} \quad (1)$$

$$= 3\lambda + \min_{\boldsymbol{x} \in X}\{(-\lambda - 2)x_1 + (1 - \lambda)x_2\} \quad (2)$$

- Interesting values of λ:
  - $\lambda \leq -2$
  - $\lambda \in [-2, 1]$
  - $\lambda \geq 1$

Solution (2)

- $\lambda \leq -2 \rightarrow \; x_1 = 0, x_2 = 0$

$$\begin{aligned} \theta(\lambda) &= 3\lambda + 0(-\lambda - 2) + 0(1 - \lambda) \\ &= 3\lambda \end{aligned}$$

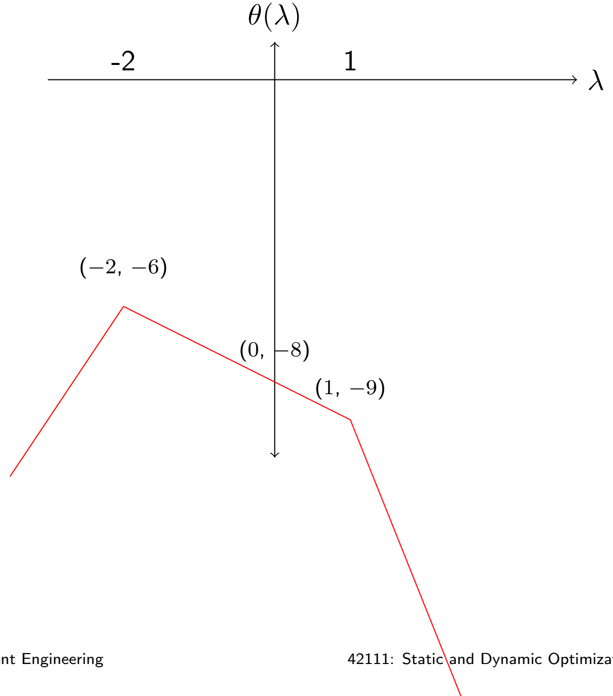
- $\lambda \in [-2, 1] \rightarrow \; x_1 = 4, x_2 = 0$

$$\begin{aligned} \theta(\lambda) &= 3\lambda + 4(-\lambda - 2) + 0(1 - \lambda) \\ &= -8 - \lambda \end{aligned}$$

- $\lambda \geq 1 \rightarrow \; x_1 = 4, x_2 = 4$

$$\begin{aligned} \theta(\lambda) &= 3\lambda + 4(-\lambda - 2) + 4(1 - \lambda) \\ &= -4 - 5\lambda \end{aligned}$$

Plot



Observations

- $\lambda^* = -2,\ \theta(\lambda^*) = -6 < f(\boldsymbol{x}^*),\ \boldsymbol{x}^* = (2, 1)$
- There is a duality gap
- $\theta(\lambda)$ does not even contain a feasible solution!

Lecture Overview

Numerical Algorithms for Optimization

- Unconstrained
- Several Variables
- Separable Programming
- Interior point methods/Barrier methods

Motivation

Numerical Algorithms for Optimization

- Nonlinear optimization is difficult, in general
- Many optimization software packages exist...
- ...but a general method that works for all problems *cannot* exist
- So: software packages will give 'wrong' answers
- Insight into algorithms: ability to *formulate* a problem with algorithm in mind

LP programs are “easy”

Two efficient algorithms

- Simplex method: insect crawling on a topaz
- Interior point methods: insect gnawing there way through a topaz.

Historical comment

- In 1983 Narendra K. Karmarkar hit the headlines in major world newspapers: he had discovered another polynomial algorithm for LP problems, that he claimed was very efficient. This type of method became known as interior point methods. Karmarkar he worked for a private company and kept the algorithm secret for years...
- This energized the scientific community to "rediscover" Karmarkars algorithm, and beat his algorithm. Experts of the simplex method fought to keep their algorithm competitive.
- The result: LP problems can be solved a million times faster than before 1983. A factor thousand is due to faster computers, another factor thousand is due to better theory.

Quality

Approximation x_0 of the point of global minimum $\hat{x}$ of a function $f : G \to \mathbb{R}$ of n variables defined on a subset G of $\mathbb{R}^n$.

Quality of a solution

- $x_0 - \hat{x}$
- $f(x_0) - f(\hat{x})$
- $\frac{f(x_0)-f(\hat{x})}{f(\tilde{x})-f(\hat{x})}$
- $|f'(x_0)|$

Quality

Quality of an algorithm

- Efficiency: each step leads to a guaranteed convergence (additional nr of decimals)
- Reliability: it provides a "certificate of quality"

Notes:

- Note: the algorithm will give an *approximation* of the optimal solution (e.g. to be ϵ distance from optimal); analytical solutions provide the exact optimal point.
- "ideal algorithms" (that are efficient and reliable) are only available for **convex** optimization problems.

One Variable Unconstrained Optimization

Let's consider the simplest case: Unconstrained optimization with just a single variable x, where the differentiable function to be minimizes is convex (or for maximization, is concave)

- The necessary and sufficient condition for a particular solution $x = x^*$ to be a global maximum is:
$$\frac{df}{dx} = 0 \qquad x = x^*$$
- Previously lectures: analytical solution methods
- What if it cannot be solved (easily) analytically?
- We can utilize search procedures to solve it numerically
- Find a sequence of trial solutions that lead towards the optimal solution

Bisection Method

- Can always be applied when $f(x)$ concave for maximization (or convex for minimization)
- It can also be used for certain other functions
- If x^* denotes the optimal solution, all that is needed is that

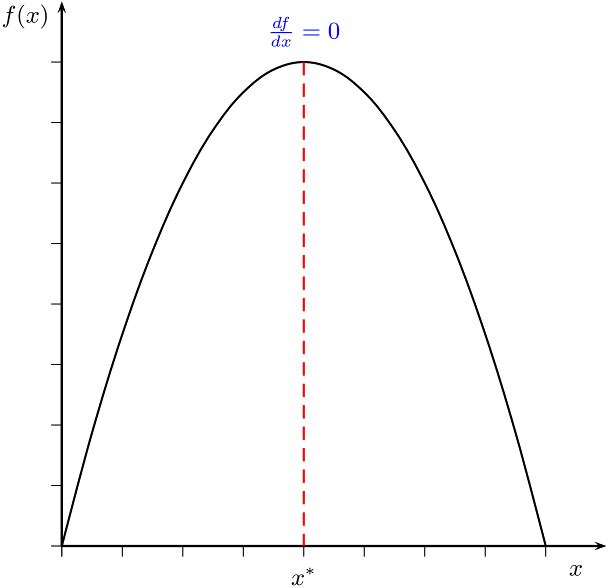
$$\frac{df}{dx} > 0 \quad \text{if } x < x^*$$

$$\frac{df}{dx} = 0 \quad \text{if } x = x^*$$

$$\frac{df}{dx} < 0 \quad \text{if } x > x^*$$

- These conditions automatically hold when $f(x)$ is concave
- The sign of the gradient indicates the direction of improvement

Example



Bisection Method

Bisection

Given two values, $\underline{x} < \overline{x}$, with $f'(\underline{x}) > 0,\ f'(\overline{x}) < 0$

- Find the midpoint $\hat{x} = \frac{\underline{x}+\overline{x}}{2}$
- Find the sign of the slope of the midpoint
- The next two values are:
  - $\overline{x} = \hat{x}$ if $f'(\hat{x}) < 0$
  - $\underline{x} = \hat{x}$ if $f'(\hat{x}) > 0$

- What is the stopping criterion?

Bisection Method

Bisection

Given two values, $\underline{x} < \overline{x}$, with $f'(\underline{x}) > 0, \ f'(\overline{x}) < 0$

- Find the midpoint $\hat{x} = \frac{\underline{x} + \overline{x}}{2}$
- Find the sign of the slope of the midpoint
- The next two values are:
    - $\overline{x} = \hat{x}$ if $f'(\hat{x}) < 0$
    - $\underline{x} = \hat{x}$ if $f'(\hat{x}) > 0$

- What is the stopping criterion?

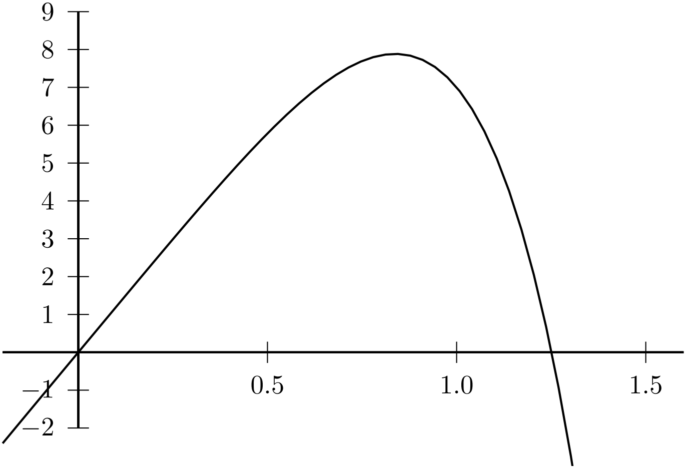
$$\overline{x} - \underline{x} < 2\epsilon$$

Bisection Method

The Problem

$$\text{maximize} \quad f(x) = 12x - 3x^4 - 2x^6$$

Bisection Method



Bisection Method

Iteration	$f'(\hat{x})$	$\underline{x}$	$\overline{x}$	$\hat{x}$	$f(\hat{x})$
0		0	2	1	7.0000
1	-12.00	0	1	0.5	5.7812
2	10.12	0.5	1	0.75	7.6948
3	4.09	0.75	1	0.875	7.8439
4	-2.19	0.75	0.875	0.8125	7.8672
5	1.31	0.8125	0.875	0.84375	7.8829
6	-0.34	0.8125	0.84375	0.828125	7.8815
7	0.51	0.828125	0.84375	0.8359375	7.8839

$$x^* \approx 0.836$$

$$0.828125 < x^* < 0.84375$$

$$f(x^*) = 7.8839$$

Bisection Method

- Intuitive and straightforward procedure
- Converges slowly
- An iteration decreases the difference between the bounds by one half
- Only information on the derivative of $f(x)$ is used
- More information could be obtained by looking at $f''(x)$

Newton Method

- Basic Idea: Approximate $f(x)$ within the neighbourhood of the current trial solution by a quadratic function
- This approximation is obtained by truncating the Taylor series after the second derivative

$$f(x_{i+1}) \approx f(x_i) + f'(x_i)(x_{i+1} - x_i) + \frac{f''(x_i)}{2}(x_{i+1} - x_i)^2$$

- Having set x_i at iteration i, this is just a quadratic function of x_{i+1}
- Can be maximized by setting its derivative to zero

Newton Method Overview

$$\max \; f(x_{i+1}) \approx f(x_i) + f'(x_i)(x_{i+1} - x_i) + \frac{f''(x_i)}{2}(x_{i+1} - x_i)^2$$

$$f'(x_{i+1}) \approx f'(x_i) + f''(x_i)(x_{i+1} - x_i)$$

$$x_{i+1} = x_i - \frac{f'(x_i)}{f''(x_i)}$$

- What is the stopping criterion?

Newton Method Overview

$$\max \; f(x_{i+1}) \approx f(x_i) + f'(x_i)(x_{i+1} - x_i) + \frac{f''(x_i)}{2}(x_{i+1} - x_i)^2$$

$$f'(x_{i+1}) \approx f'(x_i) + f''(x_i)(x_{i+1} - x_i)$$

$$x_{i+1} = x_i - \frac{f'(x_i)}{f''(x_i)}$$

- What is the stopping criterion?

$$|x_{i+1} - x_i| < \epsilon$$

Same Example

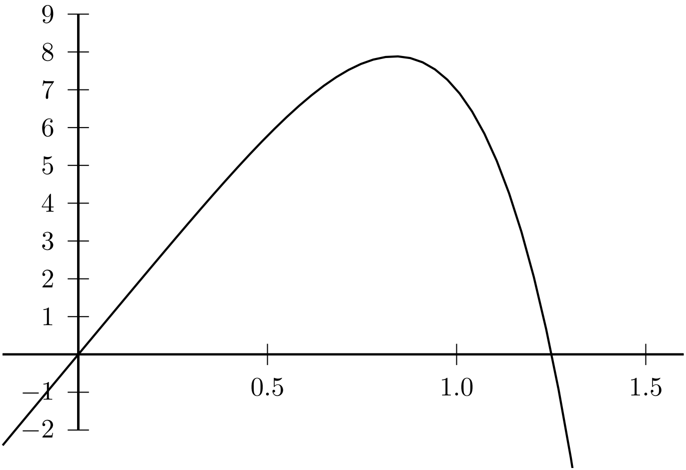
The Problem

$$\text{minimize: } f(x) = 12x - 3x^4 - 2x^6$$

$$\begin{aligned} f'(x) &= 12 - 12x^3 - 12x^5 \\ f''(x) &= -36x^3 - 60x^4 \end{aligned}$$

$$x_{i+1} = x_i + \frac{1 - x^3 - x^5}{3x^3 + 5x^4}$$

Newton Method



Newton Method

Iteration	x_i	$f(x_i)$	$f'(x_i)$	$f''(x_i)$	$f(\hat{x})$
1	1	7	-12	-96	0.875
2	0.875	7.8439	-2.1940	-62.733	0.84003
3	0.84003	7.8838	-0.1325	-55.279	0.83763
4	0.83763	7.8839	-0.0006	-54.790	0.83762

$$x^* = 083763$$

$$f(x^*) = 7.8839$$

Several variables

DTU

Newton: Multivariable

Given $\boldsymbol{x_1}$, the next iterate maximizes the quadratic approximation

$$f(\boldsymbol{x_1}) + \nabla f(\boldsymbol{x_1})(\boldsymbol{x} - \boldsymbol{x_1}) + (\boldsymbol{x} - \boldsymbol{x_1})^T H(\boldsymbol{x_1}) \frac{(\boldsymbol{x} - \boldsymbol{x_1})}{2}$$

$$\boldsymbol{x_2} = \boldsymbol{x_1} - H(\boldsymbol{x_1})^{-1} \nabla f(\boldsymbol{x_1})^T$$

Gradient search

The next iteration maximizes f along the gradient ray

$$\text{maximize: } g(t) = f(\boldsymbol{x_1} + t \nabla f(\boldsymbol{x_1})^T) \text{ s.t. } t \geq 0$$

$$\boldsymbol{x_2} = \boldsymbol{x_1} + t^* \nabla f(\boldsymbol{x_1})^T$$

Example

The Problem

$$\text{maximize: } f(x, y) = 2xy + 2y - x^2 - 2y^2$$

Example

- The vector of partial derivatives is given as

$$\nabla f(\boldsymbol{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)$$

- Here

$$\frac{\partial f}{\partial x} = 2y - 2x$$

$$\frac{\partial f}{\partial y} = 2x + 2 - 4y$$

- Suppose we select the point (x,y)=(0,0) as our initial point
- $\nabla f(0,0) = (0,2)$

Example

- Perform an iteration

$$x = 0 + t(0) = 0$$

$$y = 0 + t(2) = 2t$$

- Substituting these expressions in $f(\boldsymbol{x})$ we get

$$f(\boldsymbol{x} + t * \nabla f(\boldsymbol{x})) = f(0, 2t) = 4t - 8t^2$$

- Differentiate wrt to t

$$\frac{d}{dt}(4t - 8t^2) = 4 - 16t = 0$$

- Therefore $t^* = \frac{1}{4}$, and $\boldsymbol{x} = (0, 0) + \frac{1}{4}(0, 2) = \left(0, \frac{1}{2}\right)$

Example

- Gradient at $\boldsymbol{x} = (0, \frac{1}{2})$ is $\nabla f(0, \frac{1}{2}) = (1, 0)$
- Determine step length

$$\boldsymbol{x} = \left(0, \frac{1}{2}\right) + t(1, 0)$$

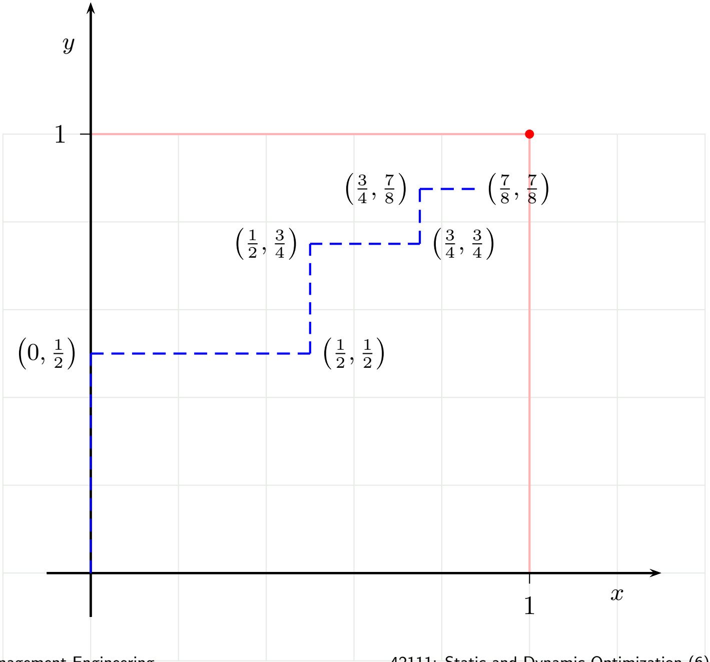
- Substituting these expressions in $f(\boldsymbol{x})$ we get

$$f(\boldsymbol{x} + t * \nabla f(\boldsymbol{x})) = f\left(t, \frac{1}{2}\right) = t - t^2 + \frac{1}{2}$$

- Differentiate wrt to t

$$\frac{d}{dt}(t - t^2 + \frac{1}{2}) = 1 - 2t = 0$$

- Therefore $t^* = \frac{1}{2}$, and $\boldsymbol{x} = \left(0, \frac{1}{2}\right) + \frac{1}{2}(1, 0) = \left(\frac{1}{2}, \frac{1}{2}\right)$



y
1
$\left(\frac{3}{4}, \frac{7}{8}\right)$
$\left(\frac{7}{8}, \frac{7}{8}\right)$
$\left(\frac{1}{2}, \frac{3}{4}\right)$
$\left(\frac{3}{4}, \frac{3}{4}\right)$
$\left(0, \frac{1}{2}\right)$
$\left(\frac{1}{2}, \frac{1}{2}\right)$
1
x

Separable Programming

The Problem

$$\begin{aligned} \text{maximize:} \quad & \textstyle\sum_j f_j(x_j) \\ \text{subject to:} \quad & A\boldsymbol{x} \leq \boldsymbol{b} \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$

- Each f_j is approximated by a piece-wise linear function

$$\begin{aligned} f(y) &= s_1 y_1 + s_2 y_2 + s_3 y_3 \\ y &= y_1 + y_2 + y_3 \\ 0 &\leq y_1 \leq u_1 \\ 0 &\leq y_2 \leq u_2 \\ 0 &\leq y_3 \leq u_3 \end{aligned}$$

Separable Programming

- Special restrictions:
  - $y_2 = 0$ whenever $y_1 < u_1$
  - $y_3 = 0$ whenever $y_2 < u_2$
- If each f_j is concave,
  - Automatically satisfied by the simplex method
- Why?

Barrier methods/Interior Point Methods

The Problem

$$\begin{aligned}\text{maximize:} \quad & f(\boldsymbol{x}) & \\ \text{subject to:} \quad & \boldsymbol{g}(\boldsymbol{x}) & \leq \boldsymbol{b} \\ & \boldsymbol{x} & \geq \boldsymbol{0}\end{aligned}$$

- For a sequence of decreasing positive r's, solve

$$\text{maximize } f(\boldsymbol{x}) - rB(\boldsymbol{x})$$

- B is a barrier function approaching ∞ as a feasible point approaches the boundary of the feasible region
- For example

$$B(\boldsymbol{x}) = \sum_i \frac{1}{b_i - g_i(\boldsymbol{x})} + \sum_j \frac{1}{x_j}$$

The Problem

$$\begin{aligned} \text{maximize:} \quad & xy \\ \text{subject to:} \quad & x^2 + y \le 3 \\ & x, y \ge 0 \end{aligned}$$

r	x	y
	1	1
1	0.90	1.36
10^{-2}	0.987	1.925
10^{-4}	0.998	1.993

- Class exercise

  Verify that the KKT conditions are satisfied at $x = 1$ & $y = 2$

Some more comments

On the power of algorithms

"General minimization schemes, such as the gradient method and the Netwon method, work well for up to four variables. Convex blackbox methods, such as the ellipsoid method, "work well" for up to 1,000 variables. Self-concordant barrier methods work well for up to 10,000 variables. A special case of self-concordant barrier methods, which is available for linear, quadratic, and semidefinite programming, the so-caled primal-dual methods, works even well for up to 1,000,000 variables."

Yurii Nesterov, *Introductory Lectures on Convex Programming: Basic Course*, Kluwer Acadamic Press, Boston, 2003.

A taste for more?

42112 Mathematical Programming Modelling (intermediate)
42114 **Integer Programming** (fundamentals)
42115 **Network Optimization** (fundamentals)
42116 Implementing OR Solution Methods (advanced)
42136 Large Scale Optimization using Decomposition (advanced)
42137 Optimization using metaheuristics (advanced)
42401 Introduction to Management Science (fundamentals)
42881 Optimisation in Public Transport (applied, advanced)
42885 Maritime Logistics (applied, advanced)
42887 Vehicle Routing and Distribution Planning (applied, advanced)

Class exercise

Separable programming

$$\begin{aligned} \text{maximize:} \quad & 32x - x^4 + 4y - y^2 \\ \text{subject to:} \quad & x^2 + y^2 \leq 9 \\ & x, y \geq 0 \end{aligned}$$

- Formulate this as an LP model using $x = 0, 1, 2, 3$ and $y = 0, 1, 2, 3$ as breakpoints for the approximating piece-wise linear functions

Evelien van der Hurk
DTU Management Engineering, Technical University of Denmark

Building 424, Room 226
evdh@dtu.dk
2800 Kgs. Lyngby, Denmark
http://www.man.dtu.dk