MS-C2105 - Introduction to Optimization
Lecture 11

Fabricio Oliveira (with modifications by Harri Hakula)

Systems Analysis Laboratory
Department of Mathematics and Systems Analysis

Aalto University
School of Science

March 22, 2022

Outline of this lecture

Newton's method for constrained problems

Newton's method for constrained problems

- Using Newton's method to solve KKT conditions

Barrier method

Primal-dual path following interior point method

Reading: Taha: Chapter 20; Winston: Chapter 11

Solving non-linear constrained optimisation problems

There are several methods available to solve nonlinear problems.

- **local solvers:** packages that employ methods that search for solutions satisfying first-order optimality conditions.
- **global solvers:** combine local solvers and specialised search methods (e.g., spatial branching).

For convex problems, local solvers can find global optimal solutions. This is a desirable feature, since local solvers are typically more efficient computationally.

We focus on a provenly efficient local solver method: barrier (or interior point) methods. They combine two central ideas:

1. The employment of Newton's method to solve optimality (KKT) conditions;
2. The use of barrier functions to eliminate inequalities.

Newton's method with equality constraints

Newton's method can (also) be used to solve systems of nonlinear equations.

- Relies on first-order approximations that are successively solved as systems of linear equations.
- Can be used as a root finding (Newton-Raphson) method to solve the system of equations arising from KKT conditions.

Newton-Raphson (NR) method: Let $f : \mathbb{R}^n \to \mathbb{R}^n$, with $f_i : \mathbb{R}^n \to \mathbb{R}$ differentiable. We wish to find x^* (a root) that solves

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}.$$

- NR starts from an initial guess x^k for x^* and iterates by finding the root x^{k+1} of a linear approximations of f at x^k.
- Under suitable conditions, the sequence $\left\{x^k\right\}$ converges to x^*.

Newton's method with equality constraints

At x^k, the first-order approximation of $f(x)$ is

$$f(x^k + d) = f(x^k) + \nabla f(x^k)^\top d,$$

where $\nabla f(x^k)$ is the Jacobian of $f(x)$ given by

$$\nabla f(x^k) = \begin{bmatrix} \nabla f_1(x^k)^\top \\ \vdots \\ \nabla f_n(x^k)^\top \end{bmatrix}.$$

We want to obtain d such that $f(x^k + d) = 0$. Therefore

$$\begin{aligned} f(x^k) + \nabla f(x^k)^\top d &= 0 \\ d &= -\nabla f(x^k)^{-1} f(x^k). \end{aligned}$$

The vector d is called Newton direction.

Newton's method with equality constraints

Algorithm Newton-Raphson method

initialise. tolerance $\epsilon > 0$, initial point x^0, iteration count $k = 0$.
while $||d|| > \epsilon$ **do**
    $d = -\nabla f(x^k)^{-1} f(x^k)$
    $x^{k+1} = x^k + d$
    $k = k + 1$
end while
return x^k

Remarks:

1. NR assumes that the Jacobian is invertible;
2. It is more efficient to solve $\nabla f(x^k)d = -f(x^k)$ using an appropriate operator than calculating inverses;
3. If x_0 is too far from optimal, NR might not converge;

Newton's method with equality constraints

Example: find the root of f with $x^0 = (1, 0, 1)$ and $\epsilon = 0.01$.

$$f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ f_3(x) \end{bmatrix} = \begin{bmatrix} x_1^2 + x_2^2 + x_3^2 - 3 \\ x_1^2 + x_2^2 - x_3 - 1 \\ x_1 + x_2 + x_3 - 3 \end{bmatrix}$$

The Jacobian is given by $\nabla f(x) = \begin{bmatrix} 2x_1 & 2x_2 & 2x_3 \\ 2x_1 & 2x_2 & -1 \\ 1 & 1 & 1 \end{bmatrix}$.

$$d^0 = -[\nabla f(x^0)]^{-1} f(x^0) = -\begin{bmatrix} 2 & 0 & 2 \\ 2 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \end{bmatrix}$$

Thus $x^1 = x^0 + d^0 = \begin{bmatrix} 3/2 & 1/2 & 1 \end{bmatrix}$. As $||x^1 - x^0|| = ||d^0|| \approx 0.7$, the method carries on until $||d^k|| < \epsilon$.

$x^* = (1, 1, 1)$ is reached after approx. 20 iterations.

Newton's method with equality constraints

NR can be employed to solve the KKT conditions of equality-constrained optimisation problems. Consider the problem

$$\begin{aligned} \text{min. } & f(x) \\ \text{s.t.: } & Ax = b \end{aligned}$$

First, consider the second-order Taylor approximation of f at x^k, where $Ax^k = b$.

$$f(x^k + \Delta x) = f(x^k) + \nabla f(x^k)^\top \Delta x + \frac{1}{2} \Delta x H(x^k) \Delta x,$$

where $H(x^k)$ is the Hessian of f at x^k and $\Delta x = x - x^k$.

The KKT conditions for the second-order approximation problem state that $x^k + \Delta x$ is optimal if exists μ such that

$$\nabla f(x^k) + H(x^k)\Delta x + A^\top \mu = 0 \tag{1}$$

$$A(x^k + \Delta x) = b \Rightarrow A\Delta x = 0 : \tag{2}$$

Using Newton's method to solve KKT conditions

These conditions are typically stated in the matrix form

$$\begin{bmatrix} H(x) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \mu \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix},$$

which is known as the Newton system.

Remarks:

- the 2^{nd}-order approximation allows for solving nonlinear optimisation problems (with linear constraints) by successively solving linear systems.
- A linearisation approach can be applied to handle nonlinear equality constraints.
- Notice that the iteration index k is omitted in this matrix form.

Using Newton's method to solve KKT conditions

Example: min. $\left\{x_1^2 - 2x_1x_2 + 4x_2^2 : 0.1x_1 - x_2 = 1\right\}$ with $x^0 = [11, 0.1]^\top$.

$\nabla f(x) = \begin{bmatrix} 2x_1 - 2x_2 \\ -2x_1 + 8x_2 \end{bmatrix}$; $H(x) = \begin{bmatrix} 2 & -2 \\ -2 & 8 \end{bmatrix}$; $A = [0.1, -1]$.

The Newton system is given by:

$$\begin{bmatrix} H(x) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \mu \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix} = \begin{bmatrix} 2 & -2 & 0.1 \\ -2 & 8 & -1 \\ 0.1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \mu \end{bmatrix} = \begin{bmatrix} -2x_1 + 2x_2 \\ 2x_1 - 8x_2 \\ 0 \end{bmatrix}.$$

For x^0, we obtain $d^1 = [\Delta x^1, \mu^1]^\top = [-11.714, -1.171, -7.142]^\top$, making $x^1 = x^0 + [-11.714, -1.171]^\top = [-0.714, -1.071]^\top$.

Remarks:

- x^1 is optimal for the problem;
- one can test for optimality by checking the KKT conditions for x^k using (1) and (2).

The barrier method

The next step towards a comprehensive optimisation framework is to deal with inequality constraints in problems of the form

$$\begin{aligned}(P) : \text{min. } & f(x) \\ \text{s.t.: } & g_i(x) \leq 0, \ i = 1, \dots, m \\ & Ax = b.\end{aligned}$$

We rely on the framework of barrier functions to represent the feasibility conditions imposed by inequality constraints.

For that, we reformulate problem P using a feasibility indicator function I that reacts to infeasibility in $g_i(x) \leq 0$, $\forall i \in \{1, \dots, m\}$.

$$\begin{aligned}\text{min. } & f(x) + \sum_{i=1}^{m} I(g_i(x)) \\ \text{s.t.: } & Ax = b,\end{aligned}$$

with $I : \mathbb{R} \to \mathbb{R}$ given by

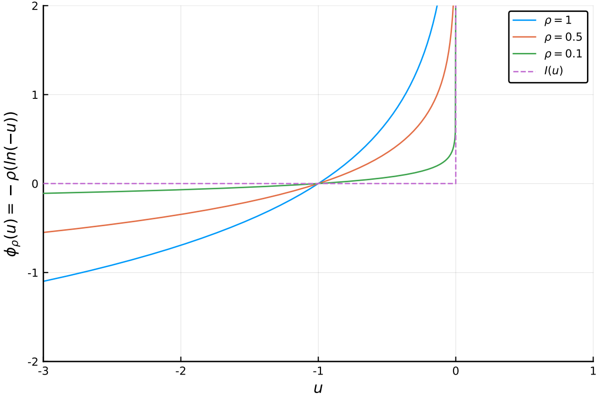
$$I(u) = \begin{cases} 0, & \text{if } u \leq 0 \\ \infty, & \text{if } u > 0 \end{cases}$$

The barrier method

To alleviate the numerical issues caused by the discontinuity of I, we approximate the indicator function using a logarithmic barrier.

$$\Phi_\rho(u) = -\rho \ln(-u)$$

where $\rho > 0$ sets the accuracy of the barrier term $\Phi_\rho(u)$.



The barrier method

Using Φ_ρ as the barrier function, the barrier problem B_ρ can be formulated as

$$(B_\rho) : \text{min. } f(x) - \rho \sum_{i=1}^{m} \ln(-g_i(x))$$
$$\text{s.t.: } Ax = b.$$

Remarks:

- Notice that the barrier problem can be solved employing NR method to its first-order optimality conditions.
- At each NR iteration, one can gradually decrease ρ by making $\rho^{k+1} = \beta\rho$ with $\beta \in (0, 1)$ (known as SUMT[1]).
- As $\rho \to 0$, $x^*(\rho) \to x^*$, where $x^*(\rho)$ and x^* are the optimal values for problems B_ρ and P, respectively.
- For small ρ, the barrier problem is challenging numerically.

[1] Sequential Unconstrained Minimisation Technique

The barrier method

Example: P : min. $\left\{f(x) = (x-3)^2 : x \geq 0\right\}$.

The barrier problem is given by

$$B_\rho : \text{min. } f(x) + \phi_\rho(x) = (x-3)^2 - \rho \ln(x)$$

The first order optimality condition for B_ρ is given by

$$\begin{aligned} & f'(x) + \phi'_\rho(x) = 0 \\ & 2(x-3) - \frac{\rho}{x} = 2x^2 - 6x - \rho = 0. \end{aligned}$$

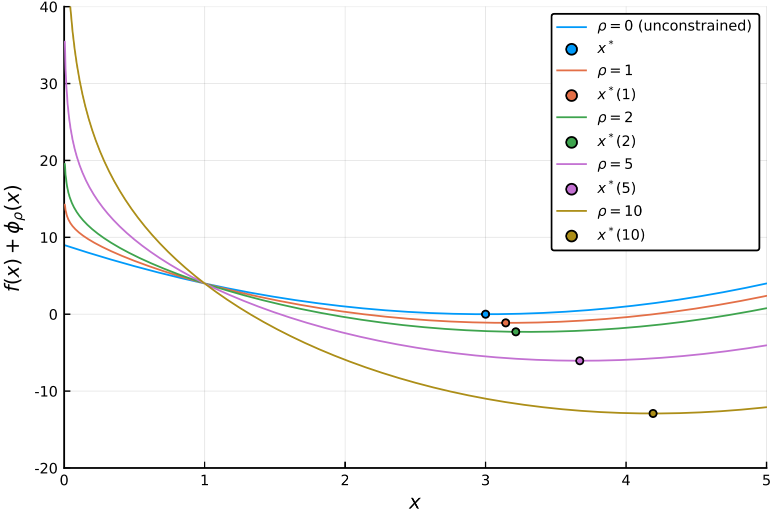
The positive solution (since $x \geq 0$) of $2x^2 - 6x - \rho = 0$ is given by

$$x^*(\rho) = \frac{6 + \sqrt{36 + 8\rho}}{4}.$$

Also, notice that $\lim_{\rho \to 0} x^*(\rho) = 3$, which is the optimal x^* for P.

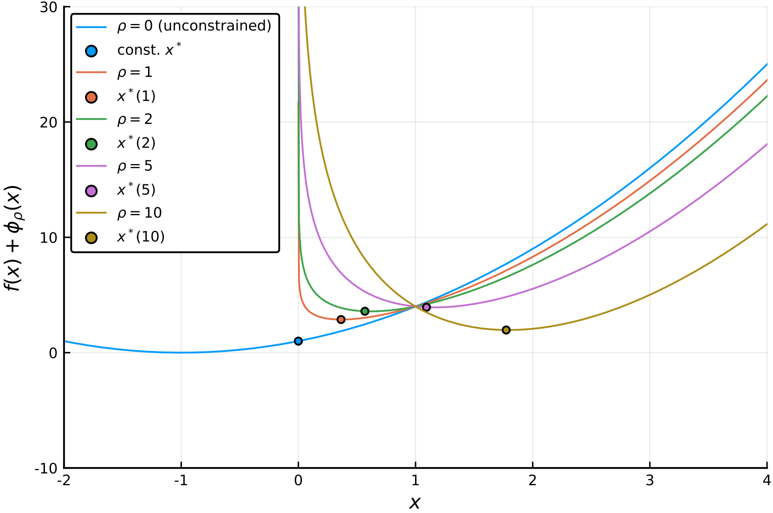
The barrier method

Example: min. $\left\{ f(x) = (x-3)^2 : x \geq 0 \right\}$



The barrier method

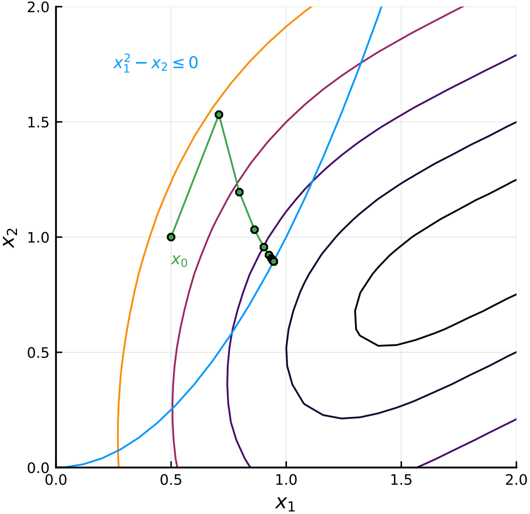
Example: min. $\left\{ f(x) = (x+1)^2 : x \geq 0 \right\}$.



The barrier method

Example:

min. $\left\{f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2 : x_1^2 - x_2 \leq 0\right\}$.



The primal-dual interior point (IP) method

The barrier framework is remarkably efficient for solving linear (or quadratic with linear constraints) optimisation problems

- differently from simplex method, it can be shown to have polynomial complexity.
- practice has shown great performance for large-scale optimisation problems.
- can be generalised of other classes of nonlinear problems.

We start with a linear problem in the standard form and formulate the barrier problem as follows.

$$(P): \text{min. } c^\top x \quad \text{s.t.: } Ax = b, \quad x \geq 0$$

$$(B_\rho): \text{min. } c^\top x - \rho \sum_{i=1}^{n} \ln(x_i) \quad \text{s.t.: } Ax = b$$

The primal-dual interior point (IP) method

Let $X = \mathbf{diag}(x)$, and e be a vector of 1's of adequate size. Thus $X^{-1} = \mathbf{diag}\left(\frac{1}{x}\right)$ and $X^{-1}e = \left(\dots \frac{1}{x_i} \dots\right)^\top$.

The KKT conditions for B_ρ can be stated as follows. First, we define the Lagrangian function

$$L(x, \mu) = c^\top x - \rho \sum_{i=1}^{n} \ln(x_i) - \mu^\top (b - Ax)$$

which leads to the following KKT (optimality) conditions:

$$\begin{aligned} \frac{\partial L(x,\mu)}{\partial x} &= c - \rho X^{-1}e - A^\top \mu = 0 \\ \frac{\partial L(x,\mu)}{\partial \mu} &= b - Ax = 0. \end{aligned}$$

Remark: notice that KKT are also sufficient for global optimality.

The primal-dual interior point (IP) method

Let $z = \rho X^{-1} e$. Then $Xz = \rho e$ or $XZe = \rho e$, with $Z = \mathbf{diag}(z)$. The KKT optimality conditions can be rewritten as

$$\begin{aligned} A^\top \mu + z &= c \\ Ax &= b \\ XZe &= \rho e. \end{aligned} \tag{3}$$

The Newton system for solving (3) using RN can be stated as

$$\begin{bmatrix} 0 & A^\top & I \\ A & 0 & 0 \\ Z & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \mu \\ \Delta z \end{bmatrix} = - \begin{bmatrix} A^\top \mu + z - c \\ Ax - b \\ XZe - \rho e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -XZe + \rho e \end{bmatrix}. \tag{4}$$

Remark: the second equality in (4) is due to primal $(Ax = b)$ and dual $(A^\top \mu + z = c)$ feasibility.

The primal-dual interior point (IP) method

Algorithm Primal-dual interior point method for LP

1: **initialise.** primal-dual feasible $w^k = (x^k, \mu^k, z^k)$, $\epsilon > 0$, ρ^k, $\beta \in (0,1)$, $k = 0$.
2: **while** $|Ax - b| > \epsilon$ and $|A^\top \mu + z - c| > \epsilon$ **do**
3: compute $\Delta w^{k+1} = (\Delta x^{k+1}, \Delta \mu^{k+1}, \Delta z^{k+1})$ using (4) and w^k.
4: $w^{k+1} = w^k + \Delta w^{k+1}$
5: $\rho^{k+1} = \beta \rho^k$, $k = k + 1$
6: **end while**
7: **return** w^k.

Remarks:

- Notice that, as $\rho \to 0$, (3) become closer to the optimality conditions for LP.
- Instead of finding optimal $x^*(\rho)$ for each ρ, the method takes a single Newton step before reducing ρ.
- Interior point methods have polynomial complexity $(O(\sqrt{n} \log \frac{1}{\epsilon}))$.

The primal-dual interior point (IP) method

Example:

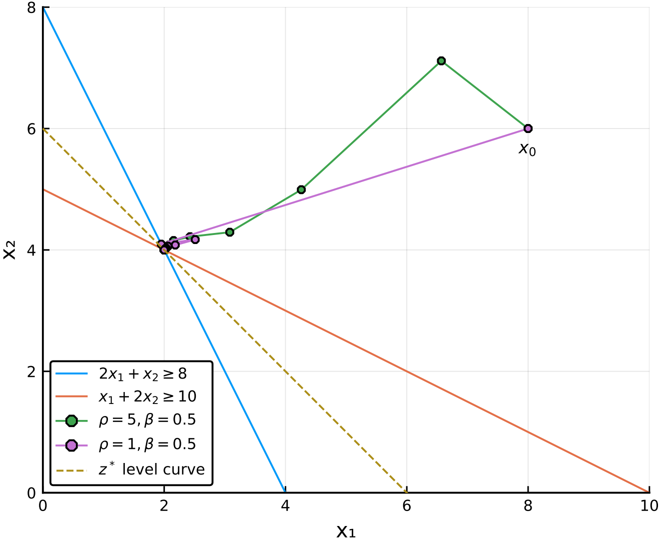
min. $\{f(x) = x_1 + x_2 : 2x_1 + x_2 \geq 8,\ x_1 + 2x_2 \geq 10,\ x_1, x_2 \geq 0\}$.

In the standard from, $A = \begin{bmatrix} 2 & 1 & -1 & 0 \\ 1 & 2 & 0 & -1 \end{bmatrix}$.

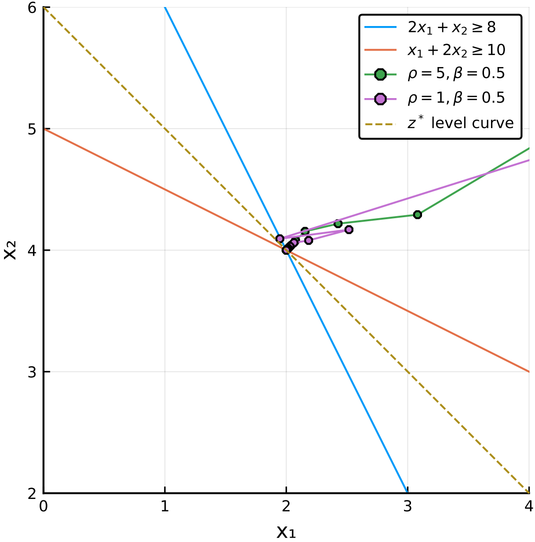
The Newton system is given by

$$\begin{bmatrix}
0 & 0 & 0 & 0 & 2 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 2 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \\
2 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
z_1^k & 0 & 0 & 0 & 0 & 0 & x_1^k & 0 & 0 & 0 \\
0 & z_2^k & 0 & 0 & 0 & 0 & 0 & x_2^k & 0 & 0 \\
0 & 0 & z_3^k & 0 & 0 & 0 & 0 & 0 & x_3^k & 0 \\
0 & 0 & 0 & z_4^k & 0 & 0 & 0 & 0 & 0 & x_4^k
\end{bmatrix}
\begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \\ \Delta x_4 \\ \Delta \mu_1 \\ \Delta \mu_2 \\ \Delta z_1 \\ \Delta z_2 \\ \Delta z_3 \\ \Delta z_4 \end{bmatrix}
=
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -x_1^k z_1^k + \rho^{k+1} \\ -x_2^k z_2^k + \rho^{k+1} \\ -x_3^k z_3^k + \rho^{k+1} \\ -x_4^k z_4^k + \rho^{k+1} \end{bmatrix}$$

The primal-dual interior point (IP) method



The barrier method



The primal-dual interior point (IP) method

Example: max. $z = x_1 + x_2 : \frac{1}{3}x_1 + x_2 \leq 5,\ \frac{1}{5}x_1 - x_2 \leq -1,$
$-\frac{8}{3}x_1 - x_2 \leq -8,\ \ \frac{1}{2}x_1 + x_2 \leq 9,\ x_1 - x_2 \leq 4, x_1, x_2 \geq 0$