This section gives an overview of different methods to solve constrained optimization problems. The general idea is to reduce the problems to unconstrained or equality-constrained optimization programs that are easier to solve.

1 Equality-constrained problems

In this section, we consider problems where the only constraints are linear equality constraints. In other words:

$$\begin{aligned} &\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \\ &\text{s.t } A\mathbf{x} = \mathbf{b} \end{aligned} \tag{P}$$

with $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. As usual, we can assume that $m \leq n$ and that A has full row rank.

1.1 Solving (P) through the dual

The dual of (P) is given by:

$$\max_{\lambda\in\mathbb{R}^m} \lambda^\intercal \mathbf{b} + \min_{\mathbf{x}\in\mathbb{R}^n} \left(f(\mathbf{x}) - \lambda^\intercal A\mathbf{x}\right)$$

This is an unconstrained concave optimization over the variable $\lambda \in \mathbb{R}^m$. In cases where the dual function is differentiable (more specifically the minimum over $\mathbf{x} \in \mathbb{R}^n$), then we can solve the dual problem using the algorithms we saw for unconstrained convex minimization and applying them to the negative of the dual function.

Once we have a solution λ^* to the dual, we can use the KKT conditions to recover an optimal solution $\mathbf{x}^* \in \mathbb{R}^n$. However, note that even if the KKT conditions fully characterize a pair of primal-dual solutions, it might not always be possible to inverse the conditions analytically and obtain a closed-form expression for $\mathbf{x}^*$ as a function of λ^*.

1.2 Solving (P) via a change of variable

The null space of A, $\ker A$ as dimension $n - m$. Let us consider a matrix $F \in \mathbb{R}^{n\times n-m}$ such that the column space of F is equal to the kernel of A. In other words, the $n - m$ columns of F form a basis of $\ker A$. Finding such an F can easily be done using Gaussian elimination.

It is now easy to see that:

$$\{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{b}\} = \{F\mathbf{z} + \hat{\mathbf{x}} \mid \mathbf{z} \in \mathbb{R}^{n-m}\}$$

where $\hat{\mathbf{x}}$ is any solution to the equation $A\mathbf{x} = \mathbf{b}$. Indeed, the two vector spaces have the same dimension $(n-m)$ and the right-hand vector space is contained in the left-hand one since:

$$A(F\mathbf{z} + \hat{\mathbf{x}}) = 0 + \mathbf{b} = \mathbf{b},\ \mathbf{z} \in \mathbb{R}^{n-m}$$

Given such a matrix F and $\hat{\mathbf{x}}$, we can now rewrite the original problem (P) as:

$$\min_{\mathbf{z} \in \mathbb{R}^{n-m}} f\big(F\mathbf{z} + \hat{\mathbf{x}}\big)$$

which is now an unconstrained convex optimization problem (the objective function is the composition of a linear and a convex function).

1.3 Equality-constrained Newton's method

Remember that for unconstrained problems, the descent direction in Newton's method $-\nabla^2 f(\mathbf{x})^{-1}\nabla f(\mathbf{x})$ was found by minimizing a quadratic approximation of f around the current solution $\mathbf{x}$:

$$\min_{\mathbf{y} \in \mathbb{R}^n} \nabla f(\mathbf{x})^\intercal \mathbf{y} + \frac{1}{2}\mathbf{y}^\intercal \nabla^2 f(\mathbf{x})\mathbf{y}$$

For the equality-constrained problem (P), the same intuition would suggest solving the following problem to find a "good" descent direction:

$$\begin{aligned} \min_{\mathbf{y} \in \mathbb{R}^n}\ & \nabla f(\mathbf{x})^\intercal \mathbf{y} + \frac{1}{2}\mathbf{y}^\intercal \nabla^2 f(\mathbf{x})\mathbf{y} \\ & A\mathbf{y} = 0 \end{aligned} \tag{1}$$

That is, we are still trying to minimize a quadratic approximation of f around $\mathbf{x}$, but under the additional constraint that the new solution $\mathbf{x} + \mathbf{y}$ stays feasible (assuming that $\mathbf{x}$ is feasible, then $A(\mathbf{x} + \mathbf{y}) = \mathbf{b}$ is equivalent to $A\mathbf{y} = 0$).

The optimal $\mathbf{y}$ solution to (1) can be found by writing the KKT conditions associated with problem (1); $\mathbf{y} \in \mathbb{R}^n$ is optimal iff there exists $\mathbf{w} \in \mathbb{R}^m$ such that:

$$\begin{aligned} \nabla f(\mathbf{x}) + \nabla^2 f(\mathbf{x})^\intercal \mathbf{y} + A^\intercal \mathbf{w} = 0 \\ A\mathbf{y} = 0 \end{aligned}$$

This is a system of $n+m$ equations in the $n+m$ variables $\mathbf{y}$ and $\mathbf{w}$. Once the optimal direction $\mathbf{y}$ is found by solving this system, then we update the current solution $\mathbf{x}$ by $\mathbf{x} \leftarrow \mathbf{x} + t\mathbf{y}$, where $t > 0$ is found either by exact or backtracking line search as already seen. Remarkably, since $A\mathbf{y} = 0$, we are guaranteed that the updated $\mathbf{x}$ remains feasible for any $t > 0$.

It is possible to analyze this adaptation of Newton's method to equality constrained problems. In fact, the analysis can be reduced to the one of unconstrained Newton's method by proving that the direction $\mathbf{y}$ defined above coincide with the Newton's descent direction for the modified unconstrained problem introduced in Section 1.2. Consequently, we obtain the same convergence guarantee as for the unconstrained Newton's method.

2 Inequality-constrained problems

We now consider the following inequality-constrained problem:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \; & f(\mathbf{x}) \\ & f_i(\mathbf{x}) \leq 0, \; 1 \leq i \leq m \end{aligned}$$

We will show how to solve this problem by reducing it to an unconstrained problem. For simplicity of exposition, we do not include equality constraints, but the analysis below can accommodate these constraints as well: in this case, the problem is reduced to an equality-constrained problem that can be solved using the techniques from Section 1.

As we saw in the section on duality (Section 7), this problem can be re-written in unconstrained form by "encoding" the constraints in the objective function:

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) + \sum_{i=1}^{m} I\big(f_i(\mathbf{x})\big)$$

where:

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

What we lose is that the objective function is now longer convex nor differentiable. The idea is then to approximate $I(u)$ by a smooth convex function. A good candidate is the so-called *logarithmic barrier function*:

$$\hat{I}_t(u) = -\frac{1}{t}\log(-u)$$

where $t > 0$ is a parameter that will be discussed below. It is interesting to note that as $t \to \infty$, $\hat{I}_t(u)$ converges pointwise to $I(u)$, which justifies the choice of $\hat{I}_t$. We will henceforth denote by $\mathbf{x}^*(t)$ the optimal solution to:

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) - \sum_{i=1}^{m} \frac{1}{t}\log\big(-f_i(\mathbf{x})\big)$$

or equivalently, writing $\phi(u) = -\log(-u)$:

$$\min_{\mathbf{x}\in\mathbb{R}^n} t f(\mathbf{x}) + \sum_{i=1}^{m} \phi\big(f_i(\mathbf{x})\big) \tag{2}$$

Note that $\mathbf{x}^*(t)$ can be found using, for example, Newton's method for unconstrained problems. The gradient and hessian of the "barrier" part of the objective are easily computed:

$$\nabla\phi\big(f_i(\mathbf{x})\big) = -\frac{\nabla f_i(\mathbf{x})}{f_i(\mathbf{x})}$$

$$\nabla^2\phi\big(f_i(\mathbf{x})\big) = \frac{\nabla f_i(\mathbf{x})\nabla f_i(\mathbf{x})^\intercal}{f_i^2(\mathbf{x})} - \frac{\nabla^2 f_i(\mathbf{x})}{f_i(\mathbf{x})}$$

Analyzing the KKT conditions of (2), it is possible to show that $\mathbf{x}^*(t)$ is $\frac{m}{t}$-optimal for any t. This suggests a very simple algorithm to solve an inequality-constrained optimization program to a given accuracy ε: solve (2) with $t = \frac{m}{\varepsilon}$. Unfortunately when ε is small, this method forces t to be very large and the Hessian of the problem (2) is ill-conditioned unless $\mathbf{x}$ is not too close to the boundary of the feasible set.

The solution is to start with a small t and solve (2) for increasing values of t, using the optimal value $\mathbf{x}^*(t)$ found at a given iteration as the starting point for the next iteration. The resulting algorithm is called the *barrier method* and is described in Algorithm 1:

Algorithm 1 Barrier method with parameter $\mu > 1$

```
t ← 1, x ← feasible solution
while m/t ≥ ε do
    compute x*(t) by solving (2) using Newton's method and using x as the initial solution
    x ← x*(t)
    t ← μt
end while
return x
```

Remark on finding a feasible initial solution. Most of the techniques we covered required starting from a feasible solution. For equality-constrained problems, this can be done by solving a linear system. For an inequality constrained problem of the following form:

$$\begin{aligned} &\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \\ &\quad f_i(\mathbf{x}) \le 0,\ 1 \le i \le m \end{aligned} \tag{3}$$

a feasible solution can be found by solving the following problem:

$$\begin{aligned} &\min_{\mathbf{x}\in\mathbb{R}^n,\, s\in\mathbb{R}} s \\ &\quad f_i(\mathbf{x}) \le s,\ 1 \le i \le m \end{aligned} \tag{4}$$

whose variables are $\mathbf{x} \in \mathbb{R}^n$ and $s \in \mathbb{R}$. It is clear that if (3) is feasible, then a solution to (4) will be a pair $(\mathbf{x}, s)$ where $s \le 0$ and $\mathbf{x}$ is feasible for (3). Problem (4) is itself an inequality-constrained program that can be solved using the barrier method. The difference is that finding a feasible starting solution for (4) is easy: consider an arbitrary $\mathbf{x} \in \mathbb{R}^n$, and choose $s = \max_{1\le i\le m} f_i(\mathbf{x})$.

References. This section is a summary of Chapters 10 and 11 from *Convex Optimization* by Boyd and Vandenberghe. More details, and in particular the convergence analysis of the various methods we saw can be found in the original book.