1 Unconstrained optimization

Let \( f : \mathbb{R}^n \to \mathbb{R} \) be convex and twice continuously differentiable and we want to solve

\[
\min f(x)
\]

**Note 1.** Since \( f \) is differentiable and convex, a necessary and sufficient condition for a point \( x^* \) to be optimal is

\[
\nabla f(x^*) = 0
\]

**Problem 1** (Quadratic minimization). Consider the problem of minimizing a quadratic function

\[
\min \frac{1}{2} x^T P x + q^T x + r
\]

where \( P \) is symmetric (but we do not assume \( P \succeq 0 \)).

a) Show that if \( P \not\succeq 0 \), i.e. the objective function is not convex, then the problem is unbounded below.

b) Write the optimality condition. Then find the optimal solution \( x^* \), when \( P \succ 0 \).

c) Write the optimality condition for the least-squares problem \( \min \|Ax - b\|_2 \), and solve it when \( A \) is injective.

d) Now suppose that \( P \succeq 0 \) (so the objective function is convex), but the optimality condition does not have a solution. Show that the problem is unbounded below.

2 Equality constrained minimization

Let \( f : \mathbb{R}^n \to \mathbb{R} \) be convex and twice continuously differentiable and \( A \in \mathbb{R}^{p \times n} \). We want to solve

\[
\begin{array}{ll}
\min & f(x) \\
\text{s.t.} & Ax = b
\end{array}
\]

**Proposition 1.** Since the problem is convex, from the KKT conditions, a point \( x^* \in \text{dom} f \) is optimal if and only if there is a \( \nu^* \in \mathbb{R}^p \) such that

\[
Ax^* = b, \quad \nabla f(x^*) + A^T \nu^* = 0 \quad \text{(called the KKT system)}
\]

**Proposition 2** (Eliminating equality constraints). Equality constraints of the form \( Ax = b \) can be eliminated by finding a particular solution \( x_0 \) of \( Ax = b \), and a matrix \( N \) whose range is the nullspace of \( A \):

\[
\{x \mid Ax = b\} = \{x_0 + r \mid Ar = 0\} = \{x_0 + Nz\}
\]
Problem 2 (Equality constrained quadratic minimization). Consider the problem

$$\min \ f(x) = \frac{1}{2} x^T P x + q^T x + r$$

s.t. \quad Ax = b$$

where $P \in \mathbb{S}_+^n$ and $A \in \mathbb{R}^{p \times n}$.

a) Write the optimality conditions.
b) Show that if the KKT system is not solvable, the quadratic optimization problem is unbounded below or infeasible.
c) Eliminate the equality constraints and write the optimality condition for the new problem.

3 Descent methods

A general descent method is an algorithm that produces a minimizing sequence $x^{(k)}$, $k = 1, 2, \cdots$, where

$$x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)}$$

$t^{(k)} > 0$ (except when $x^{(k)}$ is optimal), and $\Delta x^{(k)} \in \mathbb{R}^n$ is a descent direction such that

$$f(x^{(k+1)}) < f(x^{(k)})$$

except when $x^{(k)}$ is optimal.

Problem 3. Show that the search direction must satisfy $\nabla f(x^{(k)})^T \Delta x^{(k)} < 0$, i.e. we want an acute angle with the negative gradient.

Now we denote the update step: $x \leftarrow x + t \Delta x$. And the selection of the step size $t$ is called the line search.

Gradient descent method

A natural choice for the search direction is the negative gradient $\Delta x = -\nabla f(x)$.

Algorithm 1 Gradient descent algorithm.

```
given a starting point $x \in \text{dom } f$
    while stopping criteria is not satisfied do
        $\Delta x \leftarrow -\nabla f(x)$.
        Line search. Choose a step size $t > 0$.
        Update. $x \leftarrow x + t \Delta x$.
    end while
```

Exact line search

One line search method sometimes used in practice is exact line search, in which $t$ is chosen to minimize $f$ along the ray $\{x + t \Delta x \mid t \geq 0\}$:

$$t = \arg \min_{s \geq 0} f(x + s \Delta x)$$

An exact line search is used when the cost of minimizing problem with one variable is low compared to the cost of computing the search direction itself.
Coordinate descent method

In the coordinate descent method, line search is performed along one coordinate direction at the current point in each iteration, and different coordinate directions are used cyclically throughout the procedure. Iterations of a cycle of line search in all directions is equivalent to one gradient descent iteration. Let \{e_1, \cdots, e_n\} be some basis:

Algorithm 2 Coordinate descent algorithm.

given a starting point \(x \in \text{dom } f\).
while stopping criteria is not satified do
  for \(k = 1 \text{ to } n\) do
    Line search. Choose a step size \(t_k \in \mathbb{R}\) such that \(t_k \approx \arg\min_t f(x + te_k)\).
    Update. \(x \leftarrow x + t_k e_k\).
  end for
end while

Note that the coordinate descent method can be viewed as a descent method for which the search direction is \(\Delta x = \sum_{k=1}^n t_k e_k\), and the step size is \(t = 1\).

Problem 4. Consider the least-squares problem \(\min f(x) = \|Ax - b\|_2\), and the update step \(x(t) = x + t\Delta x\) with \(\Delta x \notin N(A)\).

a) What does \(\Delta x \notin N(A)\) imply on the search direction \(\Delta x\)?
b) Solve the exact line search.
c) Write gradient descent and coordinate descent methods for \(\min \|Ax - b\|_2\) assuming that the cost of computing the exact line search in b) is low.

Backtracking line search

In practice, most line searches are inexact: the step length is chosen to approximately solve \(\min_{s \geq 0} f(x + s\Delta s)\), or just reduce \(f\) enough. The following line search works well in practice:

Algorithm 3 Backtracking line search.

given a descent direction \(\Delta x\) for \(f\) at \(x \in \text{dom } f\), \(\alpha \in (0, 0.5)\), \(\beta \in (0, 1)\).
\(t \leftarrow 1\)
while \(f(x + t\Delta x) > f(x) + \alpha t \nabla f(x)^T \Delta x\) do
  \(t \leftarrow \beta t\)
end while