

Optimization Algorithms for A.I.

Ph.D. Program in Consumers and Markets

Curriculum: Finance, markets and regulation

Lecture notes

1 First lecture

In this course we deal with Algorithms for solving optimization problems where a single agent makes a decision on the vector $x \in \mathbb{R}^n$ in order to solve the following

$$\begin{aligned} & \underset{x}{\text{minimize}} && f(x) \\ & \text{s.t.} && Cx = c, \\ & && g_j(x) \leq b_j, \quad j = 1, \dots, m, \end{aligned} \tag{1}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $C \in \mathbb{R}^{p \times n}$, $c \in \mathbb{R}^p$ and $g_j : \mathbb{R}^n \rightarrow \mathbb{R}$. $b_j \in \mathbb{R}$ for all j . We name f the objective function and

$$X = \{x \in \mathbb{R}^n : Cx = c, g_j(x) \leq b_j, j = 1, \dots, m\}$$

the feasible set of (1). We deal with the case of continuously differentiable objective function f (see Appendix A.2 for the definition).

A vector \tilde{x} is a solution of problem (1) if

$$\tilde{x} \in X; \tag{2}$$

$$\forall x \in X \quad \text{it holds that} \quad f(\tilde{x}) \leq f(x). \tag{3}$$

Developing an algorithm for solving (1) means constructing an iterative scheme to carry out the following steps:

- choose a starting guess $x^0 \in X$;
- at each iteration k , choose a direction $d^k \in \mathbb{R}^n$;

- at each iteration k , choose a positive stepsize $\alpha^k \in \mathbb{R}$ to take in the direction d^k .

$$x^{k+1} \leftarrow x^k + \alpha^k d^k$$

During this course we will highlight how all three of the choices actually have an impact on the effectiveness of the resulting algorithm.

Let us start by considering the “easy” case of unconstrained optimization, i.e. whenever $X = \mathbb{R}^n$. Since gradient of a function f , shows the direction of fastest (local) growth, it is natural to adopt its opposite $-\nabla f$ as a direction of fastest (local) decrease for each iteration. This leads to the gradient descent method, described below where, for the moment, we adopt a fixed stepsize ($\alpha^k = \alpha$ for all k).

Algorithm 1 Gradient descent

```

1: data:  $x^0 \in \mathbb{R}^n$ ,  $\alpha > 0$ 
2: for  $k = 0, 1, \dots$  do
3:    $d^k \leftarrow -\nabla f(x^k)$ 
4:    $x^{k+1} \leftarrow x^k + \alpha d^k$ 
5: end for
```

Let us run an example in `matlab` for the quadratic function

$$f(x) = \frac{1}{2} x^\top Q x + h^\top x = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n Q_{ij} x_i x_j + \sum_{i=1}^n h_i x_i : \quad \mathbb{R}^2 \rightarrow \mathbb{R}.$$

Let us now move on and consider the constrained case, that is, $X \subset \mathbb{R}^n$. We now need to take into account that we do not want our algorithm to search outside of X . Therefore Algorithm 1 is no longer a good idea, as it ignores the existence of constraints of our problem.

To find a direction that takes into account the existence of the constraints, we rely on the **Projection operator** on X : $P_X(y)$.

The projection of a point $y \in \mathbb{R}^n$ on a set $X \subseteq \mathbb{R}^n$ is the point of X that is **closest** to y . Whenever $y \in X$, we have $P_X(y) = y$.

The projection of a point is well-defined (that is, it exists and is unique) whenever the set on which we are projecting is closed and convex (more details regarding the projection will be provided in the following lectures).

Unfortunately, solving the unconstrained minimization problem, and projecting the solution onto the feasible set X **does not**, in general, correspond to solving the constrained minimization problem.

The idea is to still use $-\nabla f$ as an indication of where to move to obtain function decrease, but to project back to the feasible set X to obtain the direction d^k .

Algorithm 2 Projected Gradient Descent

```
1: data:  $x^0 \in X$ ,  $\alpha \in (0, 1]$ 
2: for  $k = 0, 1, \dots$  do
3:    $d^k \leftarrow P_X(x^k - \nabla f(x^k)) - x^k$ 
4:    $x^{k+1} \leftarrow x^k + \alpha d^k$ 
5: end for
```

For this algorithm, we require a starting point x^0 to be feasible. On the one hand, the condition $\alpha \in (0, 1]$ is sufficient to guarantee that each iteration stays feasible when X is convex, but this condition is not necessary, and can be limiting to the algorithm speed to converge to the optimal point.

Let us run on **matlab** the same quadratic example as before, but considering $X = \{x \in \mathbb{R}^n : Ax \leq b\}$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \frac{1}{2}x^\top Qx + h^\top x \\ \text{s.t.} \quad & a_j^\top x \leq b_j, \quad j = 1, \dots, m, \end{aligned}$$

where a_j is the j^{th} row vector of A and $\frac{1}{2}x^\top Qx + h^\top x = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n Q_{ij}x_i x_j + \sum_{i=1}^n h_i x_i$ and $a_j^\top x = \sum_{i=1}^n (a_j)_i x_i$.

In order to understand the theoretical properties of the methods described, we list a few preliminary results concerning problem (1).

Theorem 1. (*Weierstrass Theorem*) *If f is continuous and X is **non-empty**, **closed** and **bounded**, then a solution \tilde{x} of problem (1) exists.*

The non-emptiness of X is a trivial assumption, but we provide a sufficient condition for the closedness of X , in the next proposition.

Proposition 2. *If g_j is continuous for every $j = 1, \dots, m$, then X is closed.*

Proof. Assume by contradiction that X is not closed. A sequence $\{x^k\} \subseteq X$ exists such that $x^k \rightarrow \bar{x} \notin X$ (Proposition 29). Therefore $\bar{j} \in \{1, \dots, m\}$ exists such that $g_{\bar{j}}(\bar{x}) > b_{\bar{j}}$, and we obtain

$$\lim_{k \rightarrow \infty} g_{\bar{j}}(x^k) \leq b_{\bar{j}} < g_{\bar{j}}(\bar{x}).$$

This implies that $g_{\bar{j}}$ is not continuous. □

Note that the equality constraints we consider are **linear**, and therefore continuous.

The closedness of X is essential both in establishing the existence of a solution and for the projection operator to be well-defined.

Aside from simple structures such as box-constraints (i.e. $X = \{x \in \mathbb{R}^n : lb \leq x \leq ub\}$, with $lb, ub \in \mathbb{R}^n$), where the boundedness of X is trivial, there is no easy condition to guarantee the boundedness of X . The next result states the existence of a solution for problem (1) by exchanging the boundedness of X with the coercivity of f . The function f is **coercive** if for every sequence $\{x^k\} \subseteq \mathbb{R}^n$ such that $\|x^k\| \rightarrow \infty$:

$$\lim_{k \rightarrow \infty} f(x^k) = \infty.$$

Theorem 3. (*Corollary of Weierstrass' theorem*) *If f is continuous and **coercive**, and X is **non-empty** and **closed**, then a solution \tilde{x} of problem (1) exists.*

We will now require the convexity of both the function f and of the set X , and we state definitions for both. More details concerning this topic can be found in the Appendix, specifically in A.2 and A.3.

A set X is said to be **convex** if for every $x, y \in X$ it holds that $[x, y] \subseteq X$, where

$$[x, y] \triangleq \{z \in \mathbb{R}^n : z = \lambda x + (1 - \lambda)y, 0 \leq \lambda \leq 1\}.$$

A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be:

- **convex** if for every $\bar{x}, y \in \mathbb{R}^n$:

$$f(\lambda \bar{x} + (1 - \lambda)y) \leq \lambda f(\bar{x}) + (1 - \lambda)f(y), \forall \lambda \in [0, 1];$$

- **strictly convex** if for every $\bar{x}, y \in \mathbb{R}^n$ with $\bar{x} \neq y$:

$$f(\lambda\bar{x} + (1 - \lambda)y) < \lambda f(\bar{x}) + (1 - \lambda)f(y), \forall \lambda \in (0, 1),$$

- **strongly convex** with modulus $\mu > 0$ if for every $\bar{x}, y \in \mathbb{R}^n$:

$$f(\lambda\bar{x} + (1 - \lambda)y) \leq \lambda f(\bar{x}) + (1 - \lambda)f(y) - \lambda(1 - \lambda)\frac{\mu}{2}\|\bar{x} - y\|^2, \forall \lambda \in [0, 1];$$

- **coercive** if for every sequence $\{x^k\} \subseteq \mathbb{R}^n$ such that $\|x^k\| \rightarrow \infty$:

$$\lim_{k \rightarrow \infty} f(x^k) = \infty.$$

In the case of a **quadratic** function

$$f(x) = \frac{1}{2}x^\top Qx + h^\top x$$

with $Q \in \mathbb{R}^{n \times n}$, $h \in \mathbb{R}^n$,

- (i) if $Q \geq 0$, then f is convex;
- (ii) if $Q > 0$, then f is strongly convex with modulus μ equal to the minimum eigenvalue of Q .

It is easy to see that strong convexity implies convexity, but the next result provides another implication that is useful to satisfy the hypotheses of Theorem 3.

Proposition 4. *Let f be strongly convex with modulus $\mu > 0$. Then f is coercive (See proof in Appendix A.2).*

The next Theorem provides sufficient conditions for problem (1) to have a unique solution.

Theorem 5. *Let f be strongly convex with modulus $\mu > 0$, and let X be convex. If \tilde{x} is a solution of problem (1), then $\tilde{x} \neq \tilde{x}$ which is a solution of problem (1) cannot exist.*

We now move on to provide a necessary first-order condition for optimality, also known as the minimum principle.

Theorem 6. *(Optimality Necessity) Let f be continuously differentiable, and let g_j be convex for every $j = 1, \dots, m$. If \tilde{x} is a solution of problem (1), then*

$$\forall x \in X \quad \text{it holds that} \quad \nabla f(\tilde{x})^\top (x - \tilde{x}) \geq 0. \quad (4)$$

The convexity of g_j is also essential to guarantee that the projection onto X is well-defined, since it guarantees the convexity of X .

Proposition 7. *If g_j is convex for every $j = 1, \dots, m$, then X is convex.*

Proof. Let $y, w \in X$, that is for every $j = 1, \dots, m$ it holds that $g_j(y) \leq b_j$ and $g_j(w) \leq b_j$. We will show that for every $\lambda \in [0, 1]$ it holds that $z = \lambda y + (1 - \lambda)w \in X$:

$$\begin{aligned} g_j(z) &= g_j(\lambda y + (1 - \lambda)w) \\ &\leq \lambda g_j(y) + (1 - \lambda)g_j(w) \\ &\leq \lambda b_j + (1 - \lambda)b_j = b_j, \end{aligned}$$

for every $j = 1, \dots, m$. This is equivalent to say that $[y, w] \subseteq X$, that is, X is convex. \square

Note that the linear equality constraints we consider are also convex.

Theorem 8. (*Optimality Sufficiency*) *Let f be continuously differentiable and convex. If $\tilde{x} \in X$ satisfies (4), then it is a solution of problem (1).*

Theorems 6 and 8 show that (4) is a necessary and sufficient condition for a feasible point to be optimal whenever (1) is a convex problem (that is, f and g_j are convex). In the forthcoming developments we are going to have the following blanket assumptions, motivated by the results of this section.

Assumptions

- To guarantee existence of solutions through Theorem 1
 - f is continuously differentiable;
 - X is nonempty and bounded;
 - g_j is continuous for all j (see Proposition 2);
- To make the minimum principle (4) a necessary (Theorem 6) and sufficient (Theorem 8) optimality condition
 - f is convex;
 - g_j is convex for all j (see Proposition 7)

2 Second lecture

Let us provide an alternative and more straightforward projected gradient algorithm, where we avoid computing the direction at each iteration and we get a more straightforward method.

Algorithm 3 Projected Gradient Descent V.2

```

1: data:  $x^0 \in X, \alpha > 0$ 
2: for  $k = 0, 1, \dots$  do
3:    $x^{k+1} \leftarrow P_X(x^k - \alpha \nabla f(x^k))$ 
4: end for

```

We have established necessary and sufficient conditions for \tilde{x} to be a solution of (1).

$$\tilde{x} \text{ is a solution of (1)} \iff \forall x \in X \text{ it holds that } \nabla f(\tilde{x})^\top (x - \tilde{x}) \geq 0.$$

We can write this condition in the form of a function that establishes the degree of optimality of a given point x :

$$\text{GAP}(x) = -\min_{y \in X} \nabla f(x)^\top (y - x).$$

Evaluating $\text{GAP}(x)$ requires solving an optimization problem with linear objective, and is therefore rather easy. The following proposition is a straightforward consequence of Theorems 6 and 8.

Proposition 9. *The following statements hold:*

- \tilde{x} is a solution of (1) $\iff \text{GAP}(\tilde{x}) = 0$;
- $\text{GAP}(x) > 0 \iff x$ is not a solution to (1).

In our convex framework we additionally have that $\text{GAP}(x)$ directly relates to the difference between the values of $f(x)$ and the optimal value $f(\tilde{x})$. This is due to the following chain

$$\begin{aligned}
f(x) - f(\tilde{x}) &\leq -\nabla f(x)^\top (\tilde{x} - x) \\
&\leq \max_{y \in X} -\nabla f(x)^\top (y - x) \\
&= -\min_{y \in X} \nabla f(x)^\top (y - x) \\
&= \text{GAP}(x),
\end{aligned}$$

where the first inequality is due to (ii) in Proposition 27 and the first equality is due to Proposition 30, both in Appendix A.4. We have therefore constructed a way to check if the iterations of our algorithm are getting closer to the optimal point. Moreover, we can check $\text{GAP}(x^k)$ and stop the algorithm whenever a satisfactory accuracy has been reached.

Let us analyze the same quadratic example as before:

$$\begin{aligned} & \underset{x}{\text{minimize}} && \frac{1}{2}x^\top Qx + h^\top x \\ & \text{s.t.} && a_j^\top x \leq b_j, \quad j = 1, \dots, m, \end{aligned}$$

where $\frac{1}{2}x^\top Qx + h^\top x = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n Q_{ij}x_i x_j + \sum_{i=1}^n h_i x_i$ and $a_j^\top x = \sum_{i=1}^n (a_j)_i x_i$.

For each iteration x^k , we can compute $\text{GAP}(x^k)$ by solving the linear problem

$$\begin{aligned} & \underset{y}{\text{minimize}} && \nabla f(x^k)^\top y - \nabla f(x^k)^\top x^k \\ & \text{s.t.} && a_j^\top y \leq b_j, \quad j = 1, \dots, m, \end{aligned}$$

where GAP is the opposite of the optimal function value and $\nabla f(x^k) = Qx^k$. This way we can check the progress of our algorithm even when we cannot visualize the iterations (for instance, when dealing with higher dimensions).

Projection Operator

Let us now provide a formal definition of the projection operator, as well as two different well-known results concerning its properties. Given $X \subseteq \mathbb{R}^n$, for each $x \in \mathbb{R}^n$, we define the projection of x on X as

$$P_X(x) \triangleq \arg \min_z \left\{ \frac{1}{2} \|z - x\|^2 : z \in X \right\}.$$

As described before, the projection of x on X is the point of X closest to x . Computing the projection requires solving a convex quadratic optimization problem in the form

$$\begin{aligned} & \underset{z}{\text{minimize}} && \frac{1}{2} z^\top \mathbb{I}_n z - x^\top z \\ & \text{s.t.} && z \in X. \end{aligned}$$

Proposition 10. *Let $x \in \mathbb{R}^n$ e $X \subseteq \mathbb{R}^n$. If X is closed and nonempty, then $P_X(x)$ exists, if X is convex, then $P_X(x)$ is unique.*

Proof. Let $f(z) = \frac{1}{2} \|z - x\|^2 = \frac{1}{2} \sum_{i=1}^n (z_i - x_i)^2$.

The function $f(z)$ is continuous and coercive, therefore Theorem 3 holds and $P_X(x)$ exists.

The function $f(z)$ is strongly convex, therefore Theorem 5 holds, and $P_X(x)$ is unique. \square

Proposition 11. *Let $x, y \in \mathbb{R}^n$ and $X \subseteq \mathbb{R}^n$ nonempty, closed and convex, the following properties hold:*

(i) $P_X(x)$ is the only point that verifies

$$(P_X(x) - x)^\top (z - P_X(x)) \geq 0, \forall z \in X$$

(ii) $\|P_X(x) - P_X(y)\| \leq \|x - y\|$.

Proof. Let $f(z) = \frac{1}{2} \|z - x\|^2 = \frac{1}{2} \sum_{i=1}^n (z_i - x_i)^2$. From Theorem 6 we have

$$(P_X(x) - x)^\top (z - P_X(x)) = \nabla f(P_X(x))^\top (z - P_X(x)) \geq 0, \forall z \in X,$$

therefore (i) holds. The uniqueness is a consequence of Theorems 8 e 5.

Using (i) we have

$$(P_X(x) - x)^\top (P_X(y) - P_X(x)) \geq 0$$

by setting $z = P_X(y)$, and

$$(P_X(y) - y)^\top (P_X(x) - P_X(y)) \geq 0$$

by writing (i) for y and setting $z = P_X(x)$. Summing these inequalities, we get

$$\begin{aligned}
0 &\leq (P_X(x) - x)^\top (P_X(y) - P_X(x)) + (P_X(y) - y)^\top (P_X(x) - P_X(y)) \\
&= (x - P_X(x))^\top (P_X(x) - P_X(y)) + (P_X(y) - y)^\top (P_X(x) - P_X(y)) \\
&= (x - P_X(x) + P_X(y) - y)^\top (P_X(x) - P_X(y)) \\
&= (x - y)^\top (P_X(x) - P_X(y)) - (P_X(x) - P_X(y))^\top (P_X(x) - P_X(y)) \\
&= (x - y)^\top (P_X(x) - P_X(y)) - \|P_X(x) - P_X(y)\|^2 \\
&\leq \|x - y\| \|P_X(x) - P_X(y)\| - \|P_X(x) - P_X(y)\|^2,
\end{aligned}$$

where the last two relations are due to the norm properties in Appendix A.1. Dividing by $\|P_X(x) - P_X(y)\|$ we get (ii) (if $\|P_X(x) - P_X(y)\| = 0$ (ii) is trivial). \square

Lipschitz Conditions

A continuous function $f : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ is Lipschitz continuous if there exists $L > 0$ such that

$$\|f(x) - f(y)\| \leq L\|x - y\|, \forall x, y \in X,$$

and we name the smallest possible L the Lipschitz constant of f . Therefore f is Lipschitz continuous if it is limited in how fast it can change: there exists a real number such that, for every pair of points on the graph of this function, the absolute value of the slope of the line connecting them is not greater than L .

Any Lipschitz continuous function is continuous.

Remark 12. *This condition can be geometrically interpreted as the existence of a double cone that can be moved along the graph so that the graph stays outside the double cone. This is clear when considering the monodimensional case $f : \mathbb{R} \rightarrow \mathbb{R}$. In this case the Lipschitz condition reduces to*

$$|f(x) - f(y)| \leq L|x - y|.$$

If we consider $x < y$, we have $|x - y| = x - y$ and the condition can be rewritten as

$$-L(x - y) + f(x) \leq f(y) \leq L(x - y) + f(x).$$

This describes the behaviour of f to the right of x , showing that it lies between two lines with slopes L and $-L$, that define the right-facing cone. The same reasoning can be applied to the case where $x > y$ to describe the left-facing cone.

A continuously differentiable function $f : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ has a Lipschitz-continuous gradient if there exists $L > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \forall x, y \in X.$$

We name the smallest possible L the Lipschitz constant of ∇f , and we write $f \in C^{1,1}$.

Proposition 13. Let $f : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ with $f \in C^{1,1}$ with constant L , then $f \in C^1$.

Proof. For any $\bar{x} \in X$ and any $\varepsilon > 0$ we have $\|x - \bar{x}\| < \varepsilon/L \Rightarrow \|\nabla f(x) - \nabla f(\bar{x})\| < \varepsilon$. \square

The following result is known as the descent lemma.

Proposition 14. Let $f : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, with $f \in C^{1,1}$ with constant L , and $\bar{x}, y \in X$, we have

$$f(y) \leq f(\bar{x}) + \nabla f(\bar{x})^\top (y - \bar{x}) + \frac{1}{2}L\|y - \bar{x}\|^2.$$

Proof. Consider the function

$$g(z) = \frac{1}{2}Lz^\top z - f(z).$$

For any $z, w \in X$ we have

$$\begin{aligned} (\nabla g(z) - \nabla g(w))^\top (z - w) &= (Lz - \nabla f(z) - Lw + \nabla f(w))^\top (z - w) \\ &= L\|z - w\|^2 - (\nabla f(z) - \nabla f(w))^\top (z - w) \\ &\geq L\|z - w\|^2 - \|\nabla f(z) - \nabla f(w)\| \|z - w\| \\ &\geq L\|z - w\|^2 - L\|z - w\|^2 = 0, \end{aligned}$$

Therefore g is convex, and we can write

$$\begin{aligned} \frac{1}{2}Ly^\top y - f(y) &= g(y) \geq g(\bar{x}) + \nabla g(\bar{x})^\top (y - \bar{x}) \\ &= \frac{1}{2}L\bar{x}^\top \bar{x} - f(\bar{x}) + (L\bar{x} - \nabla f(\bar{x}))^\top (y - \bar{x}), \end{aligned}$$

which implies

$$\begin{aligned} f(y) &\leq f(\bar{x}) + \nabla f(\bar{x})^\top (y - \bar{x}) + \frac{1}{2}Ly^\top y - \frac{1}{2}L\bar{x}^\top \bar{x} - L\bar{x}^\top (y - \bar{x}) \\ &= f(\bar{x}) + \nabla f(\bar{x})^\top (y - \bar{x}) + \frac{1}{2}L\|y - \bar{x}\|^2. \end{aligned}$$

\square

We are now ready to provide the convergence result for Algorithm 2, which is reported below.

Theorem 15. Let $f \in C^{1,1}$ with constant L , any cluster point \tilde{x} of the sequence $\{x^k\}$ produced by the Projected Gradient Algorithm 3 with $\alpha \in (0, 2/L)$ verifies

$$\tilde{x} \in X, \quad \nabla f(\tilde{x})^\top (x - \tilde{x}) \geq 0, \quad \forall x \in X.$$

Furthermore, if $\alpha \in (0, 1/L)$, we have

$$f(x^k) - f(\tilde{x}) \leq \frac{L}{2k} \|x^0 - \tilde{x}\|^2,$$

that is, the number of iterations k to ensure a precision of $f(x^k) - f(\tilde{x}) < \varepsilon$ is $\mathcal{O}(1/\varepsilon)$.

Proof. The sequence $\{x^k\}$ is well-defined due to Proposition 10. The cluster point $\tilde{x} \in X$ exists due to the boundedness and closedness of X .

Consider the generic iteration x^k . Using (i) of Proposition 11 and considering $z = x^k$, we have

$$(P_X(x^k - \alpha \nabla f(x^k)) - (x^k - \alpha \nabla f(x^k)))^\top (x^k - P_X(x^k - \alpha \nabla f(x^k))) \geq 0,$$

recalling $x^{k+1} = P_X(x^k - \alpha \nabla f(x^k))$, we have

$$(x^{k+1} - (x^k - \alpha \nabla f(x^k)))^\top (x^k - x^{k+1}) \geq 0,$$

and then

$$\nabla f(x^k)^\top (x^{k+1} - x^k) \leq -\frac{1}{\alpha} \|x^{k+1} - x^k\|^2.$$

Combining this with the inequality of Proposition 14 where we take $y = x^{k+1}$ and $\bar{x} = x^k$, we get

$$\begin{aligned} f(x^{k+1}) &\leq f(x^k) + \nabla f(x^k)^\top (x^{k+1} - x^k) + \frac{L}{2} \|x^{k+1} - x^k\|^2 \\ &\leq f(x^k) + \left(\frac{L}{2} - \frac{1}{\alpha}\right) \|x^{k+1} - x^k\|^2. \end{aligned}$$

Due to $\alpha \in (0, 2/L)$, $(\frac{L}{2} - \frac{1}{\alpha}) < 0$ and therefore the sequence $\{f(x^k)\}$ is monotone nonincreasing. By the continuity of f we have $f(x^k) \rightarrow f(\tilde{x})$ if $x^k \rightarrow \tilde{x}$, which implies $\|x^{k+1} - x^k\|^2 \rightarrow 0$. By the continuity of $g(x) = P_X(x - \alpha \nabla f(x))$ (the projection operator is continuous by (ii) of Proposition 10 and ∇f is assumed continuous) we have an infinite set of indexes $\tilde{K} \subseteq \mathbb{N}$ such that

$$0 = \lim_{k \xrightarrow{\tilde{K}} \infty} \|g(x^k) - x^k\| = \|g(\tilde{x}) - \tilde{x}\|,$$

and then $\tilde{x} = P_X(\tilde{x} - \alpha \nabla f(\tilde{x}))$. Using (i) of Proposition 11 we get

$$(\tilde{x} - \tilde{x} + \alpha \nabla f(\tilde{x}))^\top (x - \tilde{x}) \geq 0, \forall x \in X,$$

that proves the claim due to $\alpha > 0$. \square

Theorem 16. Let $f \in C^{1,1}$ with constant L strongly convex with modulus μ . The sequence $\{x^k\}$ produced by the Projected Gradient Algorithm 3 with $\alpha \in (0, 2\mu/L^2)$ converges to \tilde{x} .

Furthermore, if $\alpha = \frac{1}{L}$, we have

$$\|x^k - \tilde{x}\| \leq \tau^k \|x^0 - \tilde{x}\|,$$

with $\tau \in [0, 1]$, that is, the number of iterations k to ensure a precision of $\|x^k - \tilde{x}\| < \varepsilon$ is $\mathcal{O}(\log(1/\varepsilon))$.

Proof. For any $k \in \mathbb{N}$ we have

$$\begin{aligned}
\|x^{k+1} - \tilde{x}\|^2 &= \|P_X(x^k - \alpha \nabla f(x^k)) - P_X(\tilde{x} - \alpha \nabla f(\tilde{x}))\|^2 \\
&\leq \|x^k - \alpha \nabla f(x^k) - \tilde{x} + \alpha \nabla f(\tilde{x})\|^2 \\
&= \|x^k - \tilde{x}\|^2 - 2\alpha(x^k - \tilde{x})^T(\nabla f(x^k) - \nabla f(\tilde{x})) + \alpha^2 \|\nabla f(x^k) - \nabla f(\tilde{x})\|^2 \\
&\leq (1 - 2\alpha\mu + \alpha^2 L^2) \|x^k - \tilde{x}\|^2,
\end{aligned}$$

where in the first equality $\tilde{x} = P_X(\tilde{x} - \alpha \nabla f(\tilde{x}))$ is due to (i) of Proposition 11; the first inequality is (ii) of Proposition 11; and the last inequality holds because $f \in C^{1,1}$ and strongly convex.

Therefore by reiterating the inequality until x^0 , we have

$$\|x^k - \tilde{x}\|^2 \leq \tau^k \|x^0 - \tilde{x}\|^2,$$

with $\tau = (1 - 2\alpha\mu + \alpha^2 L^2)$.

Under the hypotheses on α , $\tau < 1$, and then

$$\lim_{k \rightarrow \infty} \|x^k - \tilde{x}\|^2 \leq \|x^0 - \tilde{x}\|^2 \lim_{k \rightarrow \infty} \tau^k = 0,$$

which implies the convergence. \square

We have seen how the Lipschitz constant of the gradient plays a key role in establishing convergence for the Projected Gradient Algorithm. Unfortunately, deriving it is not easy in general, but L is readily available in the case of quadratic functions

$$f(x) = \frac{1}{2} x^\top Q x + h^\top x,$$

where $L = \sigma_{\max}(Q)$, $\mu = \sigma_{\min}(Q)$ with $\sigma_{\max}(Q)$ and $\sigma_{\min}(Q)$ are the maximum and minimum eigenvalue of Q .

3 Third Lecture

As seen in Theorem 15, under our assumptions, we have (guaranteed) convergence for a fixed $\alpha \in (0, 2/L)$. In order to find a practical choice for the fixed stepsize we can look at the descent lemma (Proposition 14), with $y = x^{k+1}$ and $\bar{x} = x^k$:

$$f(x^{k+1}) \leq f(x^k) + \nabla f(x^k)^\top (x^{k+1} - x^k) + \frac{L}{2} \|x^{k+1} - x^k\|^2.$$

This provides us a quadratic upper bound function of x^{k+1} for the value of f at iteration $k+1$ based on the current iteration x^k . It is natural to try to minimize this upper bound w.r.t x^{k+1} , and we have:

$$\begin{aligned} & \arg \min_{x^{k+1} \in X} \left\{ f(x^k) + \nabla f(x^k)^\top (x^{k+1} - x^k) + \frac{L}{2} \|x^{k+1} - x^k\|^2 \right\} = \\ & \arg \min_{x^{k+1} \in X} \left\{ \left\| \left(x^k - \frac{1}{L} \nabla f(x^k) \right) - x^{k+1} \right\|^2 \right\} = \\ & P_X \left(x^k - \frac{1}{L} \nabla f(x^k) \right), \end{aligned}$$

which corresponds to the k^{th} iterate of Algorithm 3 with $\alpha = 1/L$. In the case of μ -strongly convex objective function, using $\alpha = 1/L$ yields

$$\|x^k - \tilde{x}\| \leq \left(1 - \frac{\mu}{L}\right)^k \|x^0 - \tilde{x}\|,$$

where \tilde{x} is the optimal solution of problem (1). Since $\mu \leq L$ for all strongly convex functions with Lipschitz continuous gradients, we see that the convergence speed depends on the value of μ/L . Recalling that for a quadratic function

$$\frac{1}{2} x^\top Q x + h^\top x$$

L is the maximum eigenvalue of Q , and μ is its minimum eigenvalue, the speed of Algorithm 3 depends on the so-called condition number of the matrix.

We can try it on the following problem:

$$\begin{aligned} & \underset{x}{\text{minimize}} \quad \frac{1}{2} x^\top Q x + h^\top x \\ & \text{s.t.} \quad \|x - c\|^2 \leq r^2, \end{aligned}$$

where $c \in \mathbb{R}^n$ and $r \in \mathbb{R}$ denote the center and radius of the hypersphere defining the feasible region X . Although $1/L$ guarantees convergence and is a good practical choice, it is by no means the best possible choice. In fact, often times it is quite slow, and a grater stepsize (that possibly does not guarantee convergence) has better results. In fact, the condition $\alpha \in (0, 2/L)$ is only sufficient, but not necessary (i.e. $\alpha \in (0, 2/L)$ implies that Algorithm 3 converges, but the opposite is not true).

Motivated by these observations, we can try to find better stepsize choices that still guarantee convergence but also provide better practical results. This means to compute a specific stepsize α^k for each iteration.

The first alternative to a fixed α^k is to consider a decreasing sequence for the stepsizes. Intuitively, taking grater steps at the first iterations (when, in general, we are furthest from the optimal solution) and then reducing the stepsize once we are closing in on the optimal point. It is rather easy to see that the sequence used for α^k must vanish as the iterations progress. However, we must be careful that it does not vanish too fast, or we risk our algorithm getting stuck far from the optimal point.

Diminishing stepsize rule $\{\alpha^k\} \subseteq \mathbb{R}$

$$\alpha^k \rightarrow 0 \quad \sum_{k=0}^{\infty} \alpha^k = \infty. \quad (5)$$

An example of a sequence that satisfies such conditions is the harmonic sequence

$$\alpha^k = \frac{1}{k^b}, \quad b \in (0, 1].$$

Therefore we have the following easy algorithm-selecting stepsize, where $\bar{\alpha}$ is the initial stepsize, to be reduced through the iterations.

Algorithm 4 Diminishing stepsize

- 1: **data:** $b \in (0, 1], \bar{\alpha} > 0$
 - 2: $\alpha^k \leftarrow \frac{\bar{\alpha}}{k^b}$
-

Methods such as the one presented above require a sequence α^k to be provided before the algorithm runs, and therefore are hard to calibrate.

The next technique is one of many so-called linesearch technique. In fact, at each iteration of the algorithm the stepsize is computed based on the local information of the function, by testing various stepsize until a specific condition is met. Let us introduce the following function, describing the iterate $k + 1$ depending on the stepsize α^k :

$$x^{k+1}(\alpha^k) = P_X(x^k - \alpha^k \nabla f(x^k)).$$

The Armijo condition we report consists in multiplying an initial stepsize $\bar{\alpha}$ by a fixed factor $\beta \in (0, 1)$ until the following condition is met:

$$f(x^{k+1}(\alpha^k)) \leq f(x^k) + \gamma \nabla f(x^k)^\top (x^{k+1}(\alpha^k) - x^k),$$

where $\gamma \in (0, 1)$. Since, for any $\alpha^k > 0$, $(x^k - x^{k+1}(\alpha^k))^\top (x^k - \alpha^k \nabla f(x^k) - x^{k+1}(\alpha^k)) \leq 0$ due to (i) in Proposition 10,

$$\nabla f(x^k)^\top (x^{k+1}(\alpha^k) - x^k) \leq -\frac{\|x^k - x^{k+1}(\alpha^k)\|}{\alpha^k} \leq 0.$$

Therefore the Armijo condition requires that f decreases sufficiently by taking a step of α^k .

Algorithm 5 Armijo stepsize rule

```

1: data:  $x^k$ ,  $\bar{\alpha} > 0$ ,  $\gamma \in (0, 1)$ ,  $\beta \in (0, 1)$ 
2:  $\alpha^k \leftarrow \bar{\alpha}$ 
3: while  $f(x^{k+1}(\alpha^k)) > f(x^k) + \gamma \nabla f(x^k)^\top (x^{k+1}(\alpha^k) - x^k)$  do
4:    $\alpha^k \leftarrow \beta \alpha^k$ 
5: end while

```

The following Proposition (whose proof can be traced back in [1, Proposition 2.3.3]) states that the procedure to compute the Armijo stepsize has finite termination.

Proposition 17. *For all $x \in X$ there exists $a(x) > 0$ such that the Armijo condition is met for all $\alpha \in (0, a(x)]$.*

The Armijo stepsize rule (Algorithm 5) allows us to pick any starting value $\bar{\alpha}$ for the stepsize, and then reduce it in case it turns out to be too big to reduce the objective function. This appealing feature comes at the cost of computing a projection for each time the stepsize is reduced (we need to compute $x^{k+1}(\alpha^k)$). Therefore a careful setting of $\bar{\alpha}$ is essential not to make the projection computations too costly.

The following is the convergence theorem for both the decreasing and the Armijo stepsize rules.

Theorem 18. *Let $f \in C^{1,1}$ with constant L , any cluster point \tilde{x} of the sequence $\{x^k\}$ produced by the Projected Gradient Algorithm 3 with α^k computed through Algorithm 4 or 5 verifies*

$$\tilde{x} \in X, \quad \nabla f(\tilde{x})^\top (x - \tilde{x}) \geq 0, \quad \forall x \in X.$$

Proof. See [1, Section 2.3]. □

We have seen in the previous section that computing the projection corresponds to solving an optimization problem. Although the Armijo stepsize shows promising practical results, the need to compute the projection at each iteration is rather expensive. Fortunately, there are some (easy) cases where the projection of a point onto X can be obtained through a closed form expression.

- **Halfspace**

Consider $X = \{x \in \mathbb{R}^n : a^\top x \leq b\}$, with $a \in \mathbb{R}^n$, $b \in \mathbb{R}$. We have

$$P_X(x) = \begin{cases} x & \text{if } a^\top x \leq b \\ x - \frac{a^\top x - b}{\|a\|^2} a & \text{if } a^\top x > b; \end{cases}$$

- **Box set**

Consider $X = \{x \in \mathbb{R}^n : l_i \leq x_i \leq u_i, i = \{1, \dots, n\}\}$ with $l_i, u_i \in \mathbb{R} \cup \{-\infty, +\infty\}$. We have

$$[P_X(x)]_i = \begin{cases} l_i & \text{if } x_i \leq l_i \\ x_i & \text{if } l_i < x_i < u_i \\ u_i & \text{if } x_i \geq u_i; \end{cases}$$

- **Hypersphere**

Consider $X = \{x \in \mathbb{R}^n : \|x - c\|^2 \leq r^2\}$, where $c \in \mathbb{R}^n$ is the center and $r \in \mathbb{R}$ is the radius. We have

$$P_X(x) = \begin{cases} x & \text{if } \|x - c\|^2 \leq r^2 \\ c + r \frac{(x-c)}{\|x-c\|} & \text{if } \|x - c\|^2 > r^2; \end{cases}$$

- **Standard simplex**

Consider $X = \{x \in \mathbb{R}^n : 0 \leq x_i \leq 1, i = \{1, \dots, n\}, \text{ and } \sum x_i = 1\}$. We have that P_X is obtained through the following finite steps method:

1. Input $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$;
2. Sort x in the ascending order as $x_{(1)} \leq \dots \leq x_{(n)}$, and set $i = n - 1$;
3. Compute $t_i = \frac{\sum_{j=i+1}^n x_{(j)} - 1}{n - i}$. If $t_i \geq x_{(i)}$ then set $\hat{t} = t_i$ and go to Step 5, otherwise set $i \leftarrow i - 1$ and redo Step 3 if $i \geq 1$ or go to Step 4 if $i = 0$;
4. Set $\hat{t} = \frac{\sum_{j=1}^n x_{(j)} - 1}{n}$;
5. Return $P_X(x) = \max\{(x - \hat{t}, 0)\}$.

- **Standard simplex with negative lower bound**

Consider $X = \{x \in \mathbb{R}^n : lb \leq x_i, i = \{1, \dots, n\}, \text{ and } \sum x_i = 1\}$, with $lb \in \mathbb{R}^n$ and $lb \leq 0$.

1. Apply the transformation

$$y = \frac{(x - lb)}{1 - \sum lb_i};$$

2. compute $P_Y(y)$ where $Y = \{x \in \mathbb{R}^n : 0 \leq x_i \leq 1, i = \{1, \dots, n\}, \text{ and } \sum x_i = 1\}$ though the previous finite-steps procedure;
3. apply the reverse transformation

$$P_X(x) = P_Y(y) \left(1 - \sum lb_i\right) + lb.$$

Lastly, consider the case of portfolio optimization, where we want to choose the portion of the budget to invest in n different assets (therefore x_i corresponds to the investment into asset i)

$$f = \delta \frac{1}{2} x^\top \Sigma x - \mu^\top x$$

and $X = \{x \in \mathbb{R}^n : 0 \leq x_i \leq 1, i = \{1, \dots, n\}, \text{ and } \sum x_i = 1\}$. The matrix Σ is the covariance matrix of the assets' returns (and is therefore always positive semidefinite) and the vector μ corresponds to the expected value of the assets return. The parameter δ regulates the degree of risk in our desired portfolio, the higher the value of δ , the more we are interested in minimizing the risk (despite the expected return).

4 Fourth Lecture

Algorithm 6 Newton method for unconstrained optimization

```

1: data:  $x^0 \in \mathbb{R}^n$ 
2: for  $k = 0, 1, \dots$  do
3:    $d^k \leftarrow$  solution of  $\nabla f(x^k) + \nabla^2 f(x^k)(d^k) = 0$ 
4:    $x^{k+1} \leftarrow x^k + d^k$ 
5: end for

```

Let us recall the optimization problem we are trying to solve.

$$\begin{aligned}
 & \underset{x}{\text{minimize}} && f(x) \\
 & \text{s.t.} && Cx = c \\
 & && g_j(x) \leq b_j, \quad j = 1, \dots, m.
 \end{aligned} \tag{6}$$

In this section we will look at a different kind of optimality condition for problem (6) (which is both necessary and sufficient just like (4)) in the case of continuously differentiable constraints g_j for $j = 1, \dots, m$.

A point \tilde{x} is said to satisfy the Karush Kun Tucker (KKT) conditions if there exist $\mu \in \mathbb{R}^p$ and $\lambda \in \mathbb{R}^m$ (called KKT multipliers) such that

$$(\text{primal}) \text{ Feasibility} \quad C\tilde{x} = c, \quad g_j(\tilde{x}) \leq b_j, \quad j = 1, \dots, m; \tag{7}$$

$$\text{Stationarity} \quad \nabla f(\tilde{x}) + C^\top \mu + \sum_{j=1}^m \lambda_j \nabla g_j(\tilde{x}) = 0; \tag{8}$$

$$(\text{dual}) \text{ Feasibility} \quad \lambda_j \geq 0, \quad j = 1, \dots, m; \tag{9}$$

$$\text{Complementarity} \quad \lambda_j(g_j(\tilde{x}) - b_j) = 0, \quad j = 1, \dots, m. \tag{10}$$

The following Theorems state that these conditions are both necessary and sufficient for the optimality of problem (6).

Theorem 19. (*Karush–Kuhn–Tucker (KKT) conditions (necessity)*) Let $f \in C^1$, and let $g_j \in C^1$ be convex for every $j = 1, \dots, m$. Assume the following Constraints Qualification (CQ) to hold

$$\exists \bar{x} : g_j(\bar{x}) < b_j, \quad j = 1, \dots, m \quad \text{and} \quad C\bar{x} = c. \tag{11}$$

If $\tilde{x} \in X$ is a solution to problem (6), then \tilde{x} satisfies the KKT conditions (7)-(10).

Proof. See [3, Theorem 1.14]. □

The condition (11) is called the Slater Constraint Qualification, and it is (usually) the easiest to verify for a given problem. There exist stronger CQs that play the same role in the proof of Theorem (19) (for more details, the reader is referred to [3, Section 1.1]).

Theorem 20. (*Karush–Kuhn–Tucker (KKT) conditions (sufficiency)*) Let $f \in C^1$, and let $g_j \in C^1$ be convex for every $j = 1, \dots, m$. If \tilde{x} satisfies the KKT conditions (7)-(10), then it is a solution to problem (6).

Proof. See [3, Theorem 1.16]. \square

Equation (10) states that whenever $g_j(\tilde{x}) < 0$, the corresponding KKT multiplier must satisfy $\lambda_j = 0$. It is easy to see that (8)-(10) can be combined to get the following

$$\nabla f(\tilde{x}) + C^\top \mu + \sum_{j \in A(\tilde{x})} \lambda_j \nabla g_j(\tilde{x}) = 0, \quad (12)$$

where $A(\tilde{x}) \triangleq \{j \in \{1, \dots, m\} : g_j(\tilde{x}) = b_j\}$ (the constraints g_j with $j \in A(\tilde{x})$ are called active constraints at \tilde{x}).

In the forthcoming developments we will assume that all requirements of Theorems 19 and 20 are satisfied. That is $f, g_j \in C^1$ and convex for all $j = 1 \dots m$ and (11) holds. The KKT multipliers μ and λ have the following interpretation. Consider the following function

$$F(b) = \min\{f(x) : g_j(x) \leq b_j, j = 1, \dots, m\},$$

which is the optimal value of f on X , depending on the constants b_j that define the constraints. It can be shown that, for all j

$$\frac{\partial F}{\partial b_j} = -\lambda_j,$$

that is, the j^{th} KKT multiplier is the improvement (since we are minimizing) of the objective function for small changes in the value of b_j defining the constraint.

Given the above necessary and sufficient conditions it is natural to try to develop a method to directly target them by addressing the stationarity equation (8).

The classical Newton method is introduced for the generic equation

$$\Phi(x) = 0, \quad (13)$$

where $\Phi(x) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a smooth mapping. Given an initial guess $x^0 \in \mathbb{R}^n$, the iteration of the Newton method is to compute x^{k+1} as the solution of the following linear equation

$$\Phi(x^k) + \Phi'(x^k)(x - x^k) = 0. \quad (14)$$

The idea is to solve the nonlinear equation (13) by iteratively addressing its linear approximation (at x^k) (14). This is computationally much simpler, and shows strong convergence properties.

Proposition 21. Let $\Phi(x) : \mathbb{R}^n \rightarrow \mathbb{R}^n \in C^1$, let \tilde{x} be a solution of (13). If $\Phi'(\tilde{x})$ is a nonsingular matrix, the iterative procedure (14) with any starting point $x^0 \in \mathbb{R}^n$ **close enough** to \tilde{x} produces a sequence $\{x^k\}$ that converges to \tilde{x} .

Let us first consider the case of unconstrained optimization (i.e. we remove all constraints of our original problem (6) therefore $X = \mathbb{R}^n$).

$$\begin{aligned} & \underset{x}{\text{minimize}} && f(x) \\ & \text{s.t.} && x \in \mathbb{R}^n. \end{aligned} \tag{15}$$

For problem (15) the necessary and sufficient KKT conditions reduce to finding \tilde{x} such that $\nabla f(\tilde{x}) = 0$. To find solutions to such equation, we can rely on the Newton method with $\Phi(x) = \nabla f(x)$, that is, given an initial guess x^0 , $x^{k+1} = x^k + d^k$, where d^k is the solution of the following linear equation

$$\nabla f(x^k) + \nabla^2 f(x^k)(d^k) = 0. \tag{16}$$

Since the Newton method requires computing the derivative of $\Phi(x) = \nabla f(x)$, we require f to be twice continuously differentiable (we write $f \in C^2$, for a formal definition see Appendix A.2).

Proposition 22. *Let $f(x) : \mathbb{R}^n \rightarrow \mathbb{R} \in C^2$, let \tilde{x} be a solution of (16). If $\nabla^2 f(\tilde{x})$ is not singular, Algorithm 6 with any starting point $x^0 \in \mathbb{R}^n$ **close enough** to \tilde{x} produces a sequence $\{x^k\}$ that converges to \tilde{x} .*

Furthermore, it holds that there exists a sequence $\{\tau_k\} \rightarrow 0$ such that

$$\|x^k - \tilde{x}\| \leq \prod_{i=0}^{k-1} \tau_i \|x^0 - \tilde{x}\|.$$

If the Hessian of f is Lipschitz continuous with constant L on $S = \{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$, that is

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L \|x - y\|$$

for all $x, y \in S$, we have

$$\|x^k - \tilde{x}\| \leq \tau^{2^k - 1} \|x^0 - \tilde{x}\|,$$

with $\tau \in (0, 1)$.

Proof. See [3, Theorem 2.15] □

Some comments regarding the previous convergence result are in order. The condition $\nabla^2 f(\tilde{x})$ non singular can be guaranteed whenever f is strongly convex on S (in fact, in this case $\nabla^2 f(x) > 0$ for any $x \in S$).

Let us now compare the convergence rates obtained for our methods so far:

given \tilde{x} a solution to (15), we have

$$\text{PGD (Algorithm 3)} \quad f(x^k) - f(\tilde{x}) \leq \frac{\|x^0 - \tilde{x}\|}{k}$$

$$\text{PGD (Algorithm 3) s.c. } f \quad \|x^k - \tilde{x}\| \leq \tau^k \|x^0 - \tilde{x}\| \quad (\tau \in (0, 1))$$

$$\text{Newton} \quad \|x^k - \tilde{x}\| \leq \prod_{i=0}^{k-1} \tau_i \|x^0 - \tilde{x}\| \quad (\tau_i \rightarrow 0)$$

$$\text{Newton } f \in C^2 \quad \|x^k - \tilde{x}\| \leq \tau^{2^k-1} \|x^0 - \tilde{x}\| \quad (\tau \in (0, 1))$$

It is evident that the fast convergence rate of Newton's method is an appealing feature, but the convergence is only guaranteed for "good" (that is, close to the optimal point) starting guesses x^0 . In order to achieve global convergence (i.e. convergence for any starting guess x^0) we need to corroborate our method with a linesearch technique such as the Armijo one. In this version, the Armijo condition is slightly different than the one used in Algorithm 5. In fact, starting from $\bar{\alpha}$, we need to multiply our step by a factor of $\beta \in (0, 1)$, until

$$f(x^k + \alpha^k d^k) \leq f(x^k) + \gamma \alpha^k \nabla f(x^k) d^k.$$

The full Newton method with linesearch is outlined in the algorithm below.

Algorithm 7 Newton method for unconstrained optimization

- 1: **data:** $x^0 \in \mathbb{R}^n$
 - 2: **for** $k = 0, 1, \dots$ **do**
 - 3: $d^k \leftarrow$ solution of $\nabla f(x^k) + \nabla^2 f(x^k)(d^k) = 0$
 - 4: compute α^k through linesearch
 - 5: $x^{k+1} \leftarrow x^k + \alpha^k d^k$
 - 6: **end for**
-

Proposition 23. *Let $f(x) : \mathbb{R}^n \rightarrow \mathbb{R} \in C^2$, let \tilde{x} be a solution of (16). If $\nabla^2 f(x) > 0$ for all x , Algorithm 7 ~~with any starting point $x^0 \in \mathbb{R}^n$~~ **close enough to \tilde{x}** produces a sequence $\{x^k\}$ that converges to \tilde{x} .*

In practice, the initial guess for the Armijo procedure is $\bar{\alpha} = 1$, since we want to preserve the fast convergence properties of the Newton method once we are close to the optimal solution.

Let us move on and consider now an optimization problem with only linear equality constraints

$$\begin{aligned} & \underset{x}{\text{minimize}} && f(x) \\ & \text{s.t.} && Cx = c, \end{aligned} \tag{17}$$

that is, $X = \{x \in \mathbb{R}^n : Cx = c\}$. For the rest of our analysis we will assume that C has full rank.

The KKT conditions for problem (17) reduce to finding x and μ such that

$$\begin{bmatrix} \nabla f(x^k) + C^\top \mu \\ Cx - c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which corresponds to finding a solution to an equation of $n + p$ variables. We can therefore apply Newton's method with $\Phi(x, \mu) : \mathbb{R}^{n+p} \rightarrow \mathbb{R}^{n+p}$ defined as

$$\Phi(x, \mu) = \begin{bmatrix} \nabla f(x^k) + C^\top \mu \\ Cx - c \end{bmatrix} = 0.$$

For each iteration k , with a feasible x^k , the Newton direction d^k for the variable x by solving the following linear equation system

$$\begin{bmatrix} \nabla f(x^k) \\ 0 \end{bmatrix} + \begin{bmatrix} \nabla^2 f(x^k) & C^\top \\ C & 0 \end{bmatrix} \begin{bmatrix} d^k \\ w^k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (18)$$

The vector w^k represents the change in the variable μ , which is not used in the algorithm. First, we note that for any feasible x^k , $x^{k+1} = x^k + \alpha^k d^k$ is always feasible for any α^k , since we have

$$C(x^k + \alpha^k d^k) = Cx^k + \alpha^k C d^k = 0$$

due to the feasibility of x^k and to the second block of p equations in (18).

The non-singularity of the KKT matrix $\begin{bmatrix} \nabla^2 f(x^k) & C^\top \\ C & 0 \end{bmatrix}$ is equivalent to the following:

$$Ax = 0, \quad x \neq 0 \implies x^\top \nabla^2 f(x) x > 0.$$

As an important special case, we note that if $\nabla^2 f(x) > 0$ for all $x \in X$, the KKT matrix must be nonsingular.

Algorithm 8 Newton method for equality-constrained optimization

- 1: **data:** $x^0 \in X$
 - 2: **for** $k = 0, 1, \dots$ **do**
 - 3: $d^k \leftarrow$ solution of (18)
 - 4: compute α^k through linesearch
 - 5: $x^{k+1} \leftarrow x^k + \alpha^k d^k$
 - 6: **end for**
-

Theorem 24. Let $f(x) : \mathbb{R}^n \rightarrow \mathbb{R} \in C^2$, let \tilde{x} be a solution of (18). Let the following conditions hold on all x in the level-set $S = \{x \in \mathbb{R}^n : f(x) \leq f(x^0), Cx = c\}$:

- $\nabla^2 f(x) \leq MI$;
- $\left\| \begin{bmatrix} \nabla^2 f(x) & C^\top \\ C & 0 \end{bmatrix}^{-1} \right\| \leq K$ for some $K > 0$;

Algorithm 8 produces a sequence $\{x^k\}$ that converges to \tilde{x} .

Proof. See [2, Section 10.2.4]. □

To introduce inequality constraints, one needs to include into the KKT conditions at x the following:

- from primal feasibility (equation (7)) $b_j - g_j(x) \geq 0$;
- from dual feasibility (equation (9)) $\lambda_j \geq 0$;
- from complementarity (equation (10)) $\lambda_j(b_j - g_j(x)) = 0$,

for all $j = 1, \dots, m$. This can be done through the so-called *complementarity functions* such as the *natural residual function* $\psi(v, z) = \min\{v, z\} : \mathbb{R}^2 \rightarrow \mathbb{R}$. In fact, such function has the property that

$$\psi(v, z) = \min\{v, z\} = 0 \iff v \geq 0, z \geq 0, vz = 0,$$

which describes the inequality-related part of the KKT conditions, considering $v = b_j - g_j(x)$ and $z = \lambda_j$ for all j . Unfortunately, any suitable complementarity function is non differentiable, and therefore requires a more complicated analysis. For more details concerning this topic, the reader is referred to [3, Section 3.2.1].

5 Fifth Lecture

We have devised a Newton method for solving equality-constrained problems, but we still have to tackle the general problem including inequality constraints.

$$\begin{aligned} & \underset{x}{\text{minimize}} && f(x) \\ & \text{s.t.} && Cx = c \\ & && g_j(x) \leq b_j, \quad j = 1, \dots, m. \end{aligned} \tag{19}$$

The first idea is to rewrite problem (19) by making the inequality constraints implicit in the objective:

$$\begin{aligned} & \underset{x}{\text{minimize}} && f(x) + \sum_{j=1}^m I_-(g_j(x) - b_j) \\ & \text{s.t.} && Cx = c, \end{aligned} \tag{20}$$

where $I_- : \mathbb{R} \rightarrow \mathbb{R}$ is called the indicator function,

$$I_-(u) = \begin{cases} 0 & \text{if } u \leq 0 \\ \infty & \text{if } u > 0. \end{cases}$$

Problem (20) has only equality constraints, but I_- is non differentiable. In order to avoid this issue, it is possible to use a **barrier** function that approximates the indicator function:

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

where $t > 0$ is a parameter that regulates the degree of approximation, the greater the value of t , the better the approximation. Substituting this into our problem we get

$$\begin{aligned} & \underset{x}{\text{minimize}} && f(x) + \frac{1}{t} \sum_{j=1}^m -\log(b_j - g_j(x)) \\ & \text{s.t.} && Cx = c. \end{aligned}$$

The barrier function is differentiable and convex, since $-\log(-u)$ is convex and monotone increasing in u , and g_j is assumed convex for all j . We define the **log barrier** function

$$\varphi(x) \triangleq \sum_{j=1}^m -\log(b_j - g_j(x)),$$

whose domain is the set of points that satisfy the inequality constraints strictly, and $\varphi \rightarrow \infty$ as $g_j(x) - b_j \rightarrow 0$ for any j . We arrive at the final (parametric) formulation of our problem equipped with the log barrier function

$$\begin{aligned} & \underset{x}{\text{minimize}} && f(x) + \frac{1}{t} \varphi(x) \\ & \text{s.t.} && Cx = c. \end{aligned} \tag{21}$$

Problem (21) is an approximation of the original problem (19) and we define $\tilde{x}(t)$ its solution. The quality of the approximation increases with the increase of parameter t , in fact, it can be shown (see [2, Section 11.2.2]) that

$$f(\tilde{x}(t)) - f(\tilde{x}) \leq \frac{m}{t},$$

where \tilde{x} is the solution of the original problem (19) and m is the number of its inequality constraints. This confirms the idea that $\tilde{x}(t) \rightarrow \tilde{x}$ as $t \rightarrow \infty$.

Based on the previous discussion one could, in theory, choose a tolerance ε , set $t = m/\varepsilon$ and solve the corresponding problem (21) and guarantee the desired error in the solution. In practice, however, the parameter t is large (i.e. when the tolerance ε is small), the objective of (21) is difficult to minimize by Newton's method, since its Hessian varies rapidly near the boundary of the (original) feasible set. This approach can therefore only be used when dealing with small problems, good starting points, and high tolerance (i.e. ε not too small).

The way to obtain stable convergence of the method is to start from a relatively low value for t , and iteratively solve problem (21) increasing t at each iteration. This is called an interior-point method, since $\tilde{x}(t)$ is strictly feasible ($g_j(\tilde{x}(t)) < b_j$) for all $t > 0$, and therefore the sequence produced by solving (21) with different values for t remains strictly feasible with respect to the original inequality constraints.

It can be shown that, whenever (21) cannot be solved through the Newton method (see Theorem 24), one can add the simple convex quadratic constraint $\|x\|^2 \leq R^2$, which ensures the hypotheses of Theorem 24 hold (see [2, Section 11.3.3]).

The scheme is as described in the following Algorithm, which requires a starting point $x^0 \in \tilde{X} = \{x \in \mathbb{R}^n : Cx = c, g_j(x) < b_j, j = 1, \dots, m\}$.

Algorithm 9 Interior-point log barrier method

```

1: data:  $x^0 \in \tilde{X}$ ,  $t^0 > 0$ ,  $\eta > 1$ 
2: for  $k = 0, 1, \dots$  do
3:   compute  $\tilde{x}(t^k)$  as a solution of  $\min f(x) + (1/t^k)\varphi(x)$  subject to  $Cx = c$ 
   starting at  $x^k$ 
4:    $x^{k+1} \leftarrow \tilde{x}(t^k)$ 
5:    $t^{k+1} \leftarrow \eta t^k$ 
6: end for
```

We refer to each execution of step 3 as an **outer iteration**, and to the Newton iterations executed at each outer iteration (for solving the subproblem with $t = t^k$) as **inner iteration**.

Some comments are in order concerning the choice of η , which involves a trade-off in the number of inner and outer iterations required. With a small value of η , each outer iteration requires solving a problem that is similar to the previous one, and therefore the previous point x^k is a good starting point for the

computation of x^{k+1} . On the other hand, a small η means more outer iterations are required to reach a desirable accuracy.

The availability of the starting point $x^0 \in \hat{X}$ is not, in general, trivial, but x^0 can be computed through the so-called phase I method, that involves solving a simple preliminary optimization problem (see [2, Section 11.4.1]).

A Definitions and Additional Results

A.1 Norms

A norm $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}$ is a function satisfying the following properties:

- (i) $\|x\| \geq 0, \forall x \in \mathbb{R}^n$
- (ii) $\|x\| = 0$ if and only if $x = 0$
- (iii) $\|ax\| = |a|\|x\|, \forall a \in \mathbb{R}, x \in \mathbb{R}^n$
- (iv) $\|x + y\| \leq \|x\| + \|y\|, \forall x, y \in \mathbb{R}^n$.

Proposition 25. *Let $x, y \in \mathbb{R}^n$, it holds that*

$$\|x - y\| \geq \|x\| - \|y\|.$$

Proof. It comes out by resorting to property (iv):

$$\|x\| = \|(x - y) + y\| \leq \|x - y\| + \|y\|.$$

□

The Euclidean norm is defined as:

$$\|x\| \triangleq \sqrt{\sum_{i=1}^n x_i^2}.$$

Proposition 26. *Let $x, y \in \mathbb{R}^n$, it holds that*

$$x^\top x = \|x\|^2, \quad |x^\top y| \leq \|x\|\|y\|.$$

A.2 Functions

A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be:

- **continuous** ($f \in C^0$) if for every $\bar{x} \in \mathbb{R}^n$ and every sequence $\{x^k\} \subseteq \mathbb{R}^n$ such that $x^k \rightarrow \bar{x}$:

$$\lim_{k \rightarrow \infty} f(x^k) - f(\bar{x}) = 0;$$

- **continuously differentiable** ($f \in C^1$) if $f \in C^0$ and $\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ exists continuous such that for every $\bar{x} \in \mathbb{R}^n$ and every sequence $\{x^k\} \subseteq \mathbb{R}^n$ such that $x^k \rightarrow \bar{x}$:

$$\lim_{k \rightarrow \infty} \frac{f(x^k) - (f(\bar{x}) + \nabla f(\bar{x})^\top (x^k - \bar{x}))}{\|x^k - \bar{x}\|} = 0;$$

- **two times continuously differentiable** ($f \in C^2$) if $f \in C^1$ and $\nabla^2 f : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ exists continuous such that for every $\bar{x} \in \mathbb{R}^n$ and every sequence $\{x^k\} \subseteq \mathbb{R}^n$ such that $x^k \rightarrow \bar{x}$:

$$\lim_{k \rightarrow \infty} \frac{f(x^k) - (f(\bar{x}) + \nabla f(\bar{x})^\top (x^k - \bar{x}) + \frac{1}{2}(x^k - \bar{x})^\top \nabla^2 f(\bar{x})(x^k - \bar{x}))}{\|x^k - \bar{x}\|^2} = 0;$$

Proposition 27. *Let $f \in C^1$, the following statements are equivalent:*

(i) *f is convex;*

(ii) *for every $\bar{x}, y \in \mathbb{R}^n$:*

$$f(y) \geq f(\bar{x}) + \nabla f(\bar{x})^\top (y - \bar{x});$$

(iii) *for every $\bar{x}, y \in \mathbb{R}^n$:*

$$(\nabla f(y) - \nabla f(\bar{x}))^\top (y - \bar{x}) \geq 0;$$

(iv) *if $f \in C^2$, for every $\bar{x} \in \mathbb{R}^n$:*

$$\nabla^2 f(\bar{x}) \geq 0,$$

that is

$$v^\top \nabla^2 f(\bar{x}) v \geq 0, \forall v \in \mathbb{R}^n.$$

Proposition 28. *Let $f \in C^1$, the following statements are equivalent:*

(i) *f is strongly convex with modulus μ ;*

(ii) *for every $\bar{x}, y \in \mathbb{R}^n$:*

$$f(y) \geq f(\bar{x}) + \nabla f(\bar{x})^\top (y - \bar{x}) + \frac{\mu}{2} \|y - \bar{x}\|^2;$$

(iii) *for every $\bar{x}, y \in \mathbb{R}^n$:*

$$(\nabla f(y) - \nabla f(\bar{x}))^\top (y - \bar{x}) \geq \mu \|y - \bar{x}\|^2;$$

(iv) *if $f \in C^2$, for every $\bar{x} \in \mathbb{R}^n$:*

$$(\nabla^2 f(\bar{x}) - \mu I) \geq 0,$$

that is

$$v^\top \nabla^2 f(\bar{x}) v \geq \mu v^\top v, \forall v \in \mathbb{R}^n.$$

Proof of Proposition 4. Let $\{x^k\} \subseteq \mathbb{R}^n$ be any sequence such that $\|x^k\| \rightarrow \infty$. By resorting to Propositions 28, 26 and 25, for every k we can write:

$$\begin{aligned}
f(x^k) &\geq f(x^0) + \nabla f(x^0)^\top (x^k - x^0) + \frac{\mu}{2} \|x^k - x^0\|^2 \\
&\geq f(x^0) - |\nabla f(x^0)^\top (x^k - x^0)| + \frac{\mu}{2} \|x^k - x^0\|^2 \\
&\geq f(x^0) - \|\nabla f(x^0)\| \|x^k - x^0\| + \frac{\mu}{2} \|x^k - x^0\|^2 \\
&= f(x^0) + \left(\frac{\mu}{2} \|x^k - x^0\| - \|\nabla f(x^0)\| \right) \|x^k - x^0\| \\
&\geq f(x^0) + \left(\frac{\mu}{2} \|x^k\| - \frac{\mu}{2} \|x^0\| - \|\nabla f(x^0)\| \right) \|x^k - x^0\|.
\end{aligned}$$

Since $\left(\frac{\mu}{2} \|x^k\| - \frac{\mu}{2} \|x^0\| - \|\nabla f(x^0)\| \right) \rightarrow \infty$ and $\|x^k - x^0\| \geq \|x^k\| - \|x^0\| \rightarrow \infty$, we obtain

$$\lim_{k \rightarrow \infty} f(x^k) \geq f(x^0) + \lim_{k \rightarrow \infty} \left(\frac{\mu}{2} \|x^k\| - \frac{\mu}{2} \|x^0\| - \|\nabla f(x^0)\| \right) \|x^k - x^0\| = \infty.$$

□

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a **quadratic** function:

$$f(x) = \frac{1}{2} x^\top Q x + h^\top x + d = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n Q_{ij} x_i x_j + \sum_{i=1}^n h_i x_i + d,$$

with $Q \in \mathbb{R}^{n \times n}$, $h \in \mathbb{R}^n$ and $d \in \mathbb{R}$, then

- (i) $f \in C^2$ and for every $x \in \mathbb{R}^n$ it holds that $\nabla f(x) = Qx + h$ and $\nabla^2 f(x) = Q$;
- (ii) if $Q \geq 0$, then f is convex;
- (iii) if $Q > 0$, then f is strongly convex with modulus μ equal to the minimum eigenvalue of Q .

A.3 Sets

Consider a set $X \subseteq \mathbb{R}^n$. We indicate with

$$\partial X \triangleq \{z \in \mathbb{R}^n : \mathcal{B}(z, \rho) \cap X \neq \emptyset, \mathcal{B}(z, \rho) \cap (\mathbb{R}^n \setminus X) \neq \emptyset, \forall \rho > 0\}$$

the **boundary** of X , where $\mathcal{B}(y, \rho)$ is the open ball with ray $\rho > 0$ and center $y \in \mathbb{R}^n$:

$$\mathcal{B}(y, \rho) \triangleq \{z \in \mathbb{R}^n : \|z - y\| < \rho\}.$$

We indicate with

$$\overset{\circ}{X} \triangleq X \setminus \partial X$$

the **interior** of X .

The set X is said to be:

- **open** if $\partial X \cap X = \emptyset$;
- **closed** if $\partial X \subseteq X$;
- **bounded** if $D > 0$ exists such that $\|x\| \leq D/2$ for every $x \in X$;
- **compact** if it is closed and bounded;
- **non-empty** if $x \in X$ exists.
- **convex** if for every $x, y \in X$ it holds that $[x, y] \subseteq X$, where

$$[x, y] \triangleq \{z \in \mathbb{R}^n : z = \lambda x + (1 - \lambda)y, 0 \leq \lambda \leq 1\}.$$

A.4 Additional Results

Proposition 29. *Let $X \subseteq \mathbb{R}^n$ be a set that is not closed. Then a sequence $\{x^k\} \subseteq X$ exists such that $x^k \rightarrow \bar{x} \notin X$.*

Proposition 30. *Every solution of the problem $\min_{x \in X} f(x)$ is a solution of $\max_{x \in X} (-f(x))$ and vice versa. Moreover, we have*

$$\min_{x \in X} f = \max_{x \in X} -f, \quad \max_{x \in X} f = \min_{x \in X} -f.$$

References

- [1] Dimitri P Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48(3):334–334, 1997.
- [2] Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [3] Alexey F Izmailov and Mikhail V Solodov. *Newton-type methods for optimization and variational problems*, volume 1. Springer, 2014.