Reading assignment: sections 4.1, 4.2 of BV; section 3.1 of the WTB.

4.1 Terminology

Standard form. The problem
\[
\begin{align*}
\min_x & \quad f_0(x) : \quad f_i(x) \leq 0, \quad i = 1, \ldots, m, \\
& \quad Ax = b, \quad i = 1, \ldots, p,
\end{align*}
\]  

(4.1)
is called a convex optimization problem if the objective function \(f_0\) is convex; the functions defining the inequality constraints \(f_i, i = 1, \ldots, m\) are convex; and \(A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p\) define the affine equality constraints. Note that, in the convex optimization model, we do not tolerate equality constraints other than affine ones.

Optimal value and feasible set. We usually denote by \(p^*\) the optimal value of the problem, and by \(\mathcal{X}\) the feasible set:
\[
\mathcal{X} = \{x \in \mathbb{R}^n : f_i(x) \leq 0, \quad i = 1, \ldots, m, \quad Ax = b\}.
\]
If \(\mathcal{X}\) is empty, then we say that the problem is not feasible. By convention, in this case we set \(p^* = +\infty\). The optimal value can also assume the value \(-\infty\), in which case we say that the problem is unbounded below. An example of a problem that is unbounded below is an unconstrained problem with \(f_0(x) = -\log x\), with domain \(\mathbb{R}_{++}\).

Feasibility problems. In some instances, we do not care about any objective function, and simply seek a feasible point. This so-called feasibility problem can be formulated in the standard form, using a zero (or constant) objective.

4.2 Optimality

Local and global optima. A feasible point \(x^* \in \mathcal{X}\) is a globally optimal (optimal for short) if \(f_0(x) = p^*\).

A feasible point \(x^* \in \mathcal{X}\) is a locally optimal if there exist \(R > 0\) such that \(f(x^*)\) equals the optimal value of problem (4.1) with the added constraint \(||x - x^*|| \leq R\). That is, \(x^*\) solves the problem “locally”.

For convex problems, any locally optimal point is globally optimal.
Indeed, let $x^*$ be a local minimizer of $f_0$ on the set $\mathcal{X}$, and let $y \in \mathcal{X}$. By definition, $x^* \in \text{dom } f_0$. We need to prove that $f_0(y) \geq f_0(x^*) = p^*$. There is nothing to prove if $f_0(y) = +\infty$, so let us assume that $y \in \text{dom } f_0$. By convexity of $f_0$ and $\mathcal{X}$, we have $x_\theta := \theta y + (1 - \theta)x^* \in \mathcal{X}$, and:

$$f_0(x_\theta) - f_0(x^*) \leq \theta(f_0(y) - f_0(x^*)�$$

Since $x^*$ is a local minimizer, the left-hand side in this inequality is nonnegative for all small enough values of $\theta > 0$. We conclude that the right hand side is nonnegative, i.e., $f_0(y) \geq f_0(x^*)$, as claimed.

Optimal set. The optimal set, $\mathcal{X}^{\text{opt}}$, is the set of optimal points. This set may be empty: for example, the feasible set may be empty. Another example is when the optimal value is only reached in the limit; think for example of the case when $n = 1$, $f_0(x) = \exp x$, and there are no constraints.

In any case, the optimal set is convex, since it can be written

$$\mathcal{X}^{\text{opt}} = \{x \in \mathbb{R}^n : f_0(x) \leq p^*, \ x \in \mathcal{X}\}.$$ 

Optimality condition. When $f_0$ is differentiable, then we know that for every $x, y \in \text{dom } f_0$,

$$f_0(y) \geq f_0(x) + \nabla f_0(x)^T(y - x).$$

Then $x$ is optimal if and only if

$$\forall y \in \mathcal{X} : \nabla f_0(x)^T(y - x) \geq 0. \quad (4.2)$$

If $\nabla f_0(x) \neq 0$, then it defines a supporting hyperplane to the feasible set at $x$.

Some examples of optimality conditions:

- For unconstrained problems, the optimality condition reduces to $\nabla f_0(x) = 0$.

- For problems with equality constraints only, the condition is that there exists a vector $\nu \in \mathbb{R}^p$ such that

$$x \in \text{dom } f_0, \ Ax = b, \ \nabla f_0(x) = A^T \nu.$$ 

Indeed, the optimality condition can be written as: $\nabla f_0(x)^T u \geq 0$ for every $u \in \mathcal{N}(A)$, which is the same as $\nabla f_0(x)^T u = 0$ for every $u \in \mathcal{N}(A)$. In turn, the latter means that $\nabla f_0(x)$ belongs to $\mathcal{R}(A^T)$, as claimed.

- For problems involving a single norm bound:

$$\min_x f_0(x) : \|x\| \leq 1,$$

the condition reads

$$x \in \text{dom } f_0, \ \|x\| \leq 1, \ -\nabla f_0(x)^T x \geq \|\nabla f_0(x)\|.\"
From this, we conclude that if the constraint is not satisfied with equality at optimum, that is, $\|x\| < 1$, then $\nabla f_0(x) = 0$, and the problem is effectively unconstrained (it has the same solution as the unconstrained problem).

The optimality conditions given above might be hard to solve. We will return to this issue later.

4.3 Equivalent problems

We can transform a convex problem into an equivalent one via a number of transformations. Sometimes the transformation is useful to obtain an explicit solution, or is done for algorithmic purposes. The transformation does not necessarily preserve the convexity properties of the problem. Here is a list of transformations that do preserve convexity.

**Epigraph form.** Sometimes it is convenient to work with the equivalent *epigraph* form:

$$
\min_{(x,t)} t : t \geq f_0(x), \ x \in \mathcal{X},
$$

in which we observe that we can always assume the cost function to be differentiable (in fact, linear), at the cost of adding one scalar variable.

**Implicit constraints.** Even though some problems appear to be unconstrained, they might contain implicit constraints. Precisely, the problem above has an implicit constraint $x \in \mathcal{D}$, where $\mathcal{D}$ is the problem’s *domain*

$$
\mathcal{D} := \text{dom} f_0 \cap \bigcap_{i=1}^{m} \text{dom} f_i.
$$

For example, the problem

$$
\min_x c^T x - \sum_{i=1}^{m} \log(b_i - a_i^T x),
$$

where $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and $a_i^T$’s are the rows of $A \in \mathbb{R}^{m \times n}$, arises as an important subproblem in some linear optimization algorithms. This problem has the implicit constraint that $x$ should belong to the interior of the polyhedron $\mathcal{P} = \{x : Ax \leq b\}$.

**Making explicit constraints implicit.** The problem in standard form can be also written in a form that makes the constraints that are explicit in the original problem, implicit. Indeed, an equivalent formulation is the unconstrained convex problem

$$
\min_x f(x)
$$
where $f$ is the sum of the original objective and the indicator function of the feasible set $\mathcal{X}$:

$$f(x) = f_0(x) + 1_{\mathcal{X}}(x) = \begin{cases} f_0(x) & x \in \mathcal{X} \\ +\infty & x \notin \mathcal{X} \end{cases}.$$ 

In the unconstrained above, the constraints are implicit. One of the main differences with the original, constrained problem is that now the objective function may not be differentiable, even if all the functions $f_i$’s are.

A less radical approach involves the convex problem with one inequality constraint

$$\min_x f_0(x) : Ax = b, \quad g(x) := \max_{1 \leq i \leq m} f_i(x) \leq 0,$$

which is equivalent to the original problem. In the above formulation, the structure of the inequality constraint is made implicit. Here, the reduction to a single constraint has a cost, since the function $g$ may not be differentiable, even though all the $f_i$’s are.

The above transformations show the versatility of the convex optimization model. They are also useful in the analysis of such problems.

**Equality constraint elimination.** We can eliminate the equality constraint $Ax = b$, by writing them as $x = x_0 + Nz$, with $x_0$ a particular solution to the equality constraint, and the columns of $N$ span the nullspace of $A$. Then we can rewrite the problem as one without equality constraints:

$$\min_z f_0(Nz + x_0) : f_i(Nz + x_0) \leq 0, \quad i = 1, \ldots, m.$$ 

This transformation preserves convexity of the the function involved. In practice, it may not be a good idea to perform this elimination. For example, if $A$ is sparse, the original problem has a sparse structure that may be exploited by some algorithms. In contrast, the reduced problem above does not inherit the sparsity characteristics, since in general the matrix $N$ is dense.

**Introducing equality constraints.** We can also introduce equality constraints in the problem. There might be several justifications for doing so: to reduce a given problem to a standard form used by off-the-shelf algorithms, or to use in decomposition methods for large-scale optimization.

The following example shows that introducing equality constraint may allow to exploit sparsity patterns inherent to the problem. Consider

$$\min_{(x_k)_{k=1}^K, y} \sum_{k=1}^K f_{0,k}(x_k) : f_k(x_k, y) \leq 0, \quad k = 1, \ldots, K.$$ 

In the above the objective involves different optimization variables, which are coupled via the presence of the “coupling” variable $y$ in each constraint. We can introduce $K$ variables
and rewrite the problem as

\[
\min_{(x_k)_{k=1}^K,(y_k)_{k=1}^K,y} \sum_{k=1}^K f_{0,k}(x_k) : f_k(x_k,y_k) \leq 0, \ y = y_k, \ k = 1,\ldots,K.
\]

Now the objective and inequality constraints are all independent (they involve different optimization variables). The only coupling constraint is now an equality constraint. This can be exploited in distributed algorithms for large-scale optimization.

**Slack variables.** Sometimes it is useful to introduce slack variables. For example, the problem with affine inequalities

\[
\min_x f_0(x) : Ax \leq b
\]
can be written

\[
\min_{x,s} f_0(x) : Ax + s = b, \ s \geq 0.
\]

**Minimizing over some variables.** We may “eliminate” some variables of the problem and reduce it to one with fewer variables. This operation preserves convexity. Specifically, if \( f \) is a convex function of the variable \( x \in \mathbb{R}^n \), and \( x \) is partitioned as \( x = (x_1,x_2) \), with \( x_i \in \mathbb{R}^{n_i}, \ i = 1,2, n = n_1 + n_2 \), then the function

\[
\tilde{f}(x_1) := \min_{x=(x_1,x_2)} f(x_1,x_2)
\]
is convex (look at the epigraph of this function). Hence the problem of minimizing \( f \) can be reduced to one involving \( x_1 \) only:

\[
\min_{x_1} \tilde{f}(x_1).
\]

The reduction may not be easy to carry out explicitly.

Here is an example where it is: consider the problem

\[
\min_{x=(x_1,x_2)} x^TQx : Ax_1 \leq b
\]
where

\[
Q := \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{pmatrix}
\]
is positive semi-definite, \( Q_{22} \) is positive-definite, and \( Q_{ij} \in \mathbb{R}^{n_i \times n_j}, \ i,j = 1,2 \). Since \( Q \succ 0 \), the above problem is convex. Furthermore, since the problem has no constraints on \( x_2 \), it is possible to solve for the minimization with respect to \( x_2 \) analytically. We end up with

\[
\min_{x_1} x_1^T\tilde{Q}x_1 : Ax_1 \leq b
\]
with $\tilde{Q} := Q_{11} - Q_{12}Q_{22}^{-1}Q_{12}^T$. From the reasoning above, we infer that $\tilde{Q}$ is positive semi-definite, since the objective function of the reduced problem is convex.

In the course of our reasoning, we have obtained an important result called the Schur Complement Lemma.

**Lemma 1 (Schur Complement).** Consider a symmetric matrix $Q$ that is decomposed into blocks:

$$Q := \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{pmatrix}.$$  

If $Q_{22} \succ 0$, then $Q \succeq 0$ if and only if $\tilde{Q} := Q_{11} - Q_{12}Q_{22}^{-1}Q_{12}^T \succ 0$. The matrix $\tilde{Q}$ is called the Schur complement of $Q_{22}$ in $Q$.

### 4.4 Problem Classes

Special classes of convex problems include:

- Linear programming (LP), where $f_0, \ldots, f_m$ in the standard form (4.1) are all affine.
- Quadratic programming (QP), which is an extension of LP in which the objective $f_0$ is now a convex quadratic function.
- Quadratically constrained quadratic programming (QCQP), which is an extension of QP in which the constraint functions $f_1, \ldots, f_m$ are convex quadratic.
- Second-order cone programming (SOCP), where $f_0$ is linear, and the constraint functions are of the form $f_i(x) = \|A_ix + b_i\|_2 + c_i^T x + d_i$, for appropriate data $A_i, b_i, c_i, d_i$.
- Semi-definite programming (SDP), where $f_0$ is linear, and the constraint functions are of the form $f_i(x) = \lambda_{\text{max}}(A_i^0 + x_1A_i^1 + \ldots + x_nA_i^n)$, and where the matrices $A_i^j$ are all symmetric. Here, $\lambda_{\text{max}}$ denotes the largest eigenvalue of its (symmetric) matrix argument.

All these problems are in a hierarchy, $SDP \supset SOCP \supset QCQP \supset QP \supset LP$. They are often described as “conic programs”, for reasons explained in lecture 5. The software CVX relies on conic programming.

Numerous examples of conic problems are given in the Web Textbook, sections 3.2-3.6.

### Exercises

1. Consider the problem of projection on an affine subspace:

$$\min_x \|x\|_2 : Ax = b.$$  

Here, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are given.
(a) Under which condition is the problem feasible? From now on, assume that is the case.

(b) Express the solution to the problem using linear algebra tools.

(c) Solve the problem using the optimality condition (4.2).

2. State and prove a Schur complement lemma that covers the case when the block $Q_{22}$ is only positive semi-definite.

3. Robust least-squares with interval coefficient matrix. An interval matrix in $\mathbb{R}^{m \times n}$ is a matrix whose entries are intervals:

$$
A = \{ A \in \mathbb{R}^{m \times n} \mid |A_{ij} - \bar{A}_{ij}| \leq R_{ij}, \ i = 1, \ldots, m, \ j = 1, \ldots, n \}.
$$

The matrix $\bar{A} \in \mathbb{R}^{m \times n}$ is called the nominal value or center value, and $R \in \mathbb{R}^{m \times n}$, which is elementwise nonnegative, is called the radius.

The robust least-squares problem, with interval matrix, is

$$
\text{minimize} \ \sup_{A \in A} \|Ax - b\|_2,
$$

with optimization variable $x \in \mathbb{R}^n$. The problem data are $A$ (i.e., $\bar{A}$ and $R$) and $b \in \mathbb{R}^m$. The objective, as a function of $x$, is called the worst-case residual norm. The robust least-squares problem is evidently a convex optimization problem.

(a) Formulate the interval matrix robust least-squares problem as a standard optimization problem, e.g., a QP, SOCP, or SDP. You can introduce new variables if needed. Your reformulation should have a number of variables and constraints that grows linearly with $m$ and $n$, and not exponentially.

(b) Consider the specific problem instance with $m = 4$, $n = 3$,

$$
A = \begin{bmatrix}
60 \pm 0.05 & 45 \pm 0.05 & -8 \pm 0.05 \\
90 \pm 0.05 & 30 \pm 0.05 & -30 \pm 0.05 \\
0 \pm 0.05 & -8 \pm 0.05 & -4 \pm 0.05 \\
30 \pm 0.05 & 10 \pm 0.05 & -10 \pm 0.05
\end{bmatrix}, \quad b = \begin{bmatrix}
-6 \\
-3 \\
18 \\
-9
\end{bmatrix}.
$$

(The first part of each entry in $A$ gives $\bar{A}_{ij}$; the second gives $R_{ij}$, which are all 0.05 here.) Find the solution $x_{\text{ls}}$ of the nominal problem (i.e., minimize $\|\bar{A}x - b\|_2$), and robust least-squares solution $x_{\text{rls}}$. For each of these, find the nominal residual norm, and also the worst-case residual norm. Make sure the results make sense.