1 Problem transformations

A convex problem can often be transformed into an equivalent one which is more amenable to solution (e.g from nondifferentiable to differentiable, even linear).

A device which is often used is the introduction of new variables (so called "slack variables"). The number of Newton steps of interior point methods does not increase much if the problem dimension is increased and the complexity of the equation solving does not increase much either if the new matrices are sparse and the sparsity is used by the algorithm. Below are two examples that show what is possible:

Example 1.1. L^1-norm.
The L^1-norm $f(x) = \|x\|_1 = \sum_j |x_j|$ is plainly nondifferentiable. Suppose we want to minimize $f(x)$ given some constraints Cts. Then we can introduce additional variables t_j and minimize

$$\tilde{f}(x,t) = \sum\nolimits_j t_j$$

subject to the constraints Cts and additional constraints

$$t \geq 0 \quad \text{and} \quad -t \leq x \leq t. \tag{1}$$

In similar fashion a constraint of the form $\|x\|_1 \leq C$ can be replaced with $\sum_j t_j \leq C$ and (1).

Example 1.2. $|\cdot|_L$-norm.
A more interesting example is the following: fix a positive integer $L \leq n$. For $x \in \mathbb{R}^n$ define the *L-largest norm* $|x|_L$ as follows: order the absolute values $|x_j|$ in increasing order and let $|x|_L$ be the sum of the L-largest of these.

This is easily seen to define a norm on $\mathbb{R}^n$. This norm is relevant for example if we want to minimize the sum of the L largest residuals $|(Ax)_i - b_i|$ in a general linear equation $Ax = b$ without exact solution.

Note that if we minimize $|Ax - b|_L$ instead of $\|Ax - b\|_1$, we will be able to push the sum of the L largest residuals down more at the expense of an increase in the remaining residuals which however, by definition are smaller!

We claim that the problem

$$\text{minimize } |x|_L \quad \text{subject to constraints } Cts \tag{2}$$

is equivalent to the following problem: introduce new variables t of the same dimension as x (one t_j for each x_j) and one new scalar variable ρ and minimize the new objective function

$$\hat{f}(x,t,\rho) = L\rho + \sum\nolimits_j t_j \tag{3}$$

subject to the constraints Cts augmented with the constraints $t, \rho \geq 0$ and $-t_j - \rho \leq x_j \leq t_j + \rho$, that is, $|x_j| \leq t_j + \rho$.

Note first that for all x, t, ρ satisfying these constraints we trivially have $\hat{f}(x, t, \rho) \geq |x|_L$.

Conversely, for any $x \in \mathbb{R}^n$ let ρ be the $L+1$-largest of the $|x_j|$ and set

$$t_j := \begin{cases} 0 & \text{if } |x_j| \leq \rho \\ |x_j| - \rho & \text{if } |x_j| > \rho \end{cases}$$

Plainly then ρ and t satisfy the new constraints and $t_j = 0$ if and only if $|x_j|$ is not among the L largest absolute values $|x_k|$. It follows that

$$\hat{f}(x, t, \rho) = L\rho + \sum\nolimits_j t_j = |x|_L$$

The claim follows easily from this. In similar fashion a constraint of the form $|x|_L \leq C$ is replaced with a linear constraint $L\rho + \sum_j t_j \leq C$.

Example 1.3. *Maximum of convex functions.*
Suppose we have r convex functions $f_j(x)$, $j = 1, \ldots, r$ and a set of constraints C. We want to minimize the maximum

$$g(x) = \max\{\, f_1(x), \ldots, f_n(x)\,\}$$

subject to these same constraints. The function $g(x)$ is again convex but it is not in general differentiable even if the f_j are. We can bypass this difficulty by introducing one additional variable y and passing to the equivalent problem

$$? = argmin\, h(x, y) := y \quad \text{subject to } f_j(x) \leq y,\ j = 1, \ldots, r \text{ and } C.$$

Note that this means that we can likewise maximize the minimum of concave functions g_j (by simply replacing each g_j with $-g_j$).

Example 1.4. *Sum of the L largest of a number of convex functions*
We have seen in example 1.2 how we can deal with the sum of the absolute values $|x_j|$ of the L-largest of the components of the independent variable x. The same principle can be applied to any number of convex functions f_j.

Suppose we are given r convex functions $f_j(x) \geq 0$, $j = 1, \ldots, r$ and a positive integer $L \leq r$. Define the function $g(x)$ to be the sum of the L largest values among $f_1(x), \ldots, f_n(x)$. Then g is again convex. Given any set of convex constraints C we can reformulate the problem

$$? = argmin\, g(x) \quad \text{subject to } C \tag{4}$$

as follows: introduce one new slack variable ρ and for each function f_j a new slack variable t_j, new constraints

$$t_j, \rho \geq 0 \text{ and } f_j(x) \leq t_j + \rho,\ j = 1, \ldots, r, \tag{5}$$

and a new objective function

$$\hat{f}(x,t,\rho) := L\rho + \sum\nolimits_j t_j \tag{6}$$

We claim that the problem (4) is equivalent with the new problem

$$? = \mathit{argmin}\,\hat{f}(x,t,\rho) \quad \text{subject to } C \text{ and (5).} \tag{7}$$

To see this consider any x and recall that by assumption $f_j(x) \geq 0$. Let $\rho := f_k(x)$, where $f_k(x)$ is the $(L{+}1)$-st largest among the values $f_j(x)$, $j = 1, 2, \ldots, r$ and $t_j := (f_j(x) - \rho)_+$. Then the point (x, t, ρ) satisfies the new constraints and we have

$$g(x) = L\rho + \sum\nolimits_j t_j = \hat{f}(x,t,\rho).$$

To see this note that if $f_i(x)$ is among the L largest of the $f_j(x)$, then $t_i = f_i(x) - \rho$, i.e. $f_i(x) = t_i + \rho$ while for all other $f_j(x)$ we have $t_j = 0$. In addition we also have

$$g(x) \leq \hat{f}(x,t,\rho)$$

for all points (x, t, ρ) satisfying the new constraints. It follows that $f(x,t,\rho)$ assumes *the same minimum value* as $g(x)$ at a point (x_0, t_0, ρ_0) such that $g(x)$ assumes the minimum at $x = x_0$ subject to the original constraints C (which do not affect the new variables t_j and ρ).

Example 1.5. *Simplifying the objective function*
Let us start with some simple observations: obviously the generic problem

$$? = \mathit{argmin}\, f(x) \quad \text{subject to constraints } C \tag{8}$$

is equivalent to the following problem with one new variable t

$$? = \mathit{argmin}\,\hat{f}(x,t) := t \quad \text{subject to } f(x) \leq t \text{ and } C. \tag{9}$$

Here the new objective function $\hat{f}(x,t) = t$ is *linear* but we have not gained anything since we have simply pushed the problem into the new constraint $f(x) \leq t$.

Another trivial observation is the following: if the function $h : \mathbb{R} \to \mathbb{R}$ is increasing, then minimizing the function $h(f(x))$ is equivalent to minimizing $f(x)$ (under any constraints). If h is not increasing we might try to replace the problem

$$? = \mathit{argmin}\, h(f(x)) \quad \text{subject to constraints } C$$

with the equivalent problem

$$? = \mathit{argmin}\, t \quad \text{subject to } h(f(x)) \leq t \text{ and } C$$

and then solve the inequality $h(u) \leq t$ for $u \in \mathbb{R}$ which often has a solution of the form $a_j(t) \leq u \leq b_j(t)$, $j = 1, \ldots, r$, where these inequalities are combined by logical OR (only one of them needs to hold). We could then pass to an equivalent sequence of problems

$$? = argmin\, \hat{f}(x,t) := t \quad \text{subject to } a_j(t) \leq f(x) \leq b_j(t) \text{ and } C,$$

where we add the inequalities $a_j(t) \leq f(x) \leq b_j(t)$, $j = 1, \ldots, r$, one by one. We solve these problems and pick the best of the minimizers obtained in these problems. The problem here is that in general the constraints $a_j(t) \leq f(x)$ and $f(x) \leq b_j(t)$ are not admissible (not convex in the variable (x,t)). Rewriting them as

$$a_j(t) - f(x) \leq 0 \quad \text{and} \quad f(x) - b_j(t) \leq 0$$

we see that we need $f(x)$ to be both convex and concave, i.e. affine, and $a_j(t)$ has to be convex while $b_j(t)$ has to be concave.

If these conditions are met, the new problems are convex even though the original objective function $h(f(x)$ may not have been convex at all!

The function $h(u) = u^2$ provides an example where the inequality $h(u) \leq t$ leads to $a(t) := -\sqrt{t} \leq u \leq b(t) := \sqrt{t}$ where $a(t)$ and $b(t)$ have the required properties.