CHEM-E7225 Advanced Process Control D Winter 2026

Lecture 04 – Nonlinear Programming (B)

Otacilio "Minho" Neto (otacilio.bneto@pm.me)

In the previous lecture, we studied the basics of nonlinear programming, and discussed some cool classes of problems. This lecture continues our journey by addressing the question: how to solve a nonlinear program?

1 Preface: Solving unconstrained optimization problems

Introductory calculus courses teach us that the extremum points (i.e., minima, maxima, and saddle points) of a function $f : \mathbb{R}^{N_x} \to \mathbb{R}$ are the points $x \in \mathbb{R}^{N_x}$ at which the gradient vanishes, $\nabla f(x) = 0$. Consequently, the problem

$$\underset{x \in \mathbb{R}^{N_x}}{\text{minimize}} \quad f_0(x) \tag{1}$$

must have its solutions $x^\star \in \Omega$ *necessarily* satisfying $\nabla f_0(x^\star) = 0$. This condition is sometimes called the *stationary condition* (or property) of extremum points. If f_0 is a convex function, the stationary condition is a *necessary and sufficient* condition for optimality, that is, for characterizing its global minima (see Lecture Notes L03, §2.1). Hence, we can use this fact to reformulate Problem (1) as the *root-finding problem*

$$\text{Find } x^\star \in \mathbb{R}^{N_x} \text{ such that } \nabla f_0(x^\star) = 0. \tag{2}$$

The Problem (2) is something that (in principle) can be solved easily, sometimes even in closed-form.

Example 1.1 ***Solving an unconstrained quadratic optimization problem***

An unconstrained convex quadratic program is equivalently reformulated as the root-finding problem

$$\underset{x \in \mathbb{R}^{N_x}}{\text{minimize}} \quad (1/2)x^\mathsf{T} P x + q^\mathsf{T} x + r \qquad \Longleftrightarrow \qquad \text{Find } x^\star \in \mathbb{R}^{N_x} \text{ such that } Px^\star + q = 0, \tag{3}$$

from which we can obtain the analytical solution $x^\star = -P^\dagger q$ (which is unique if P is positive-definite).

In general, the roots of $\nabla f_0(\cdot)$ cannot be obtained in closed-form. *Root-finding algorithms* are numerical methods to (approximately) compute them. An extremely important such algorithm is *Newton's method*, which uses derivative information to efficiently compute the zeroes of continuous functions. It became specially important in nonlinear programming since, as we discuss later, it can exploit the *structure* that often arises in such problems.

Remark. *If f_0 is not convex, the stationary condition is* necessary *but not* sufficient*; meaning that there are $x \in \mathbb{R}^{N_x}$ satisfying $\nabla f_0(x) = 0$ which are not solutions to Problem (1). These include maxima, saddle points, and local minima.*

2 Solving nonlinear programs

The practice of solving a mathematical problem, computationally, corresponds to coming up with an equation that characterizes its solution, then formulate an equivalent root-finding problem to find it. In optimization, these system of equations are called *optimality conditions*; they include the stationary condition mentioned above.

In a general nonlinear program,

$$\begin{aligned} \underset{x \in \mathbb{R}^{N_x}}{\text{minimize}} \quad & f_0(x) && \text{(4a)} \\ \text{subject to} \quad & f_i(x) = 0, \quad i = 1, \dots, N_f && \text{(4b)} \\ & h_i(x) \leq 0, \quad i = 1, \dots, N_h && \text{(4c)} \end{aligned}$$

the stationary condition by itself does not lead to an equivalent root-finding problem due to the equality and inequality constraints. We have to deal with them, somehow. The standard procedure to formulate optimality conditions is using the principle of *Lagrangian duality*, which we (briefly) overview in the following.

2.1 Lagrangian duality

Let's proceed in a constructive manner. Since we know how to solve unconstrained problems, we might want to start by getting rid of the constraints using indicator functions. Thus, we rewrite Problem (4) as

$$\underset{x\in\mathbb{R}^{N_x}}{\text{minimize}} \quad f_0(x) + \textstyle\sum_{i=1}^{N_f} I_{\{0\}}(f_i(x)) + \sum_{i=1}^{N_h} I_{\mathbb{R}_{\leq 0}}(h_i(x)). \tag{5}$$

Unfortunately, the augmented objective function above is discontinuous (that is, it does not have a gradient). Instead, let's consider linear functions instead of indicator functions, to obtain the unconstrained optimization problem

$$\underset{x\in\mathbb{R}^{N_x}}{\text{minimize}} \quad \underbrace{f_0(x) + \textstyle\sum_{i=1}^{N_f} \lambda_i f_i(x) + \sum_{i=1}^{N_h} \nu_i h_i(x)}_{L(x,\lambda,\nu)} \tag{6}$$

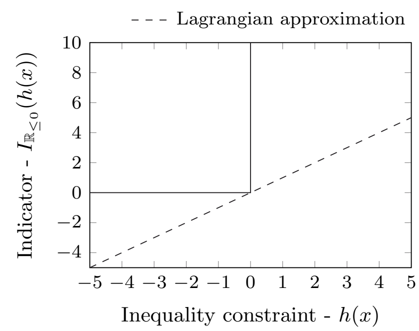

Figure 1. Incorporating constraints to f_0.

given the *multipliers* (or *slack variables*) $\lambda \in \mathbb{R}^{N_f}$ and $\nu \in \mathbb{R}^{N_f}$, with $\nu \succeq 0$. This approximation is interpreted as allowing constraint violations, but augmenting the objective to penalize them, as depicted in Figure 1. The augmented objective $L(\cdot)$ in known as the *Lagrangian* associated with Problem (4), and the optimal value $g(\lambda,\mu) = \min_x L(x,\lambda,\mu)$ is known as the *dual function* associated with it. Interestingly, the dual function is always a concave function.

Clearly, this is a poor approximation to the original problem. However, it has a special property: the dual function provides a lower bound on the solution of Problem (4), that is, we have $g(\lambda,\mu) \leq p^\star$ for any pair of multipliers (λ,ν). In particular, we can obtain the *best lower bound* as $d^\star = \max_{\lambda,\nu}\{g(\lambda,\nu) \mid \nu \succ 0\}$, that is, by solving a concave maximization problem. Thus, $d^\star \leq p^\star$. This is always true, even for nonconvex programs.

Example 2.1 ***Dual problem of standard-form LPs***

The best lower bound on the solution of a LP (the 'primal') can be obtained from another LP (the 'dual'),

$$\begin{array}{rl} \text{minimize} & c^\mathsf{T} x \\ \text{subject to} & Ax - b = 0 \\ & x \succeq 0 \end{array} \quad \underset{\text{Primal problem}}{\overset{\text{Dual problem}}{\rightleftharpoons}} \quad \begin{array}{rl} \text{maximize} & -b^\mathsf{T}\lambda \\ \text{subject to} & A^\mathsf{T}\lambda - \nu + c = 0 \\ & \nu \succeq 0 \end{array} \tag{7}$$

Karush-Kuhn-Tucker (KKT) optimality conditions

Now, let's restrict ourselves to convex problems. If the feasible set has an *interior point* (a qualification known as *Slater's condition* [1, §5]), then we say that *strong duality* holds and $d^\star = p^\star$. In other words, the optimal value $p^\star$ can be obtained by minimizing the Lagrangian with the *optimal multipliers*, $L(x,\lambda^\star,\nu^\star)$. Similarly, the optimal value $d^\star$ can be obtained by minimizing the Lagrangian with the optimal solution, $L(x^\star,\lambda,\nu)$. Thus, we can obtain the *primal optimal* $x^\star$ and *dual optimal* $(\lambda^\star,\nu^\star)$ points by considering the stationarity of $L(\cdot)$ with respect to each argument, simultaneously. The result is the famous *Karush-Kuhn-Tucker (KKT) optimality conditions*, in vector form,

$$\nabla f_0(x^\star) + \nabla f(x^\star)^\mathsf{T}\lambda^\star + \nabla h(x^\star)^\mathsf{T}\nu^\star = 0, \tag{8a}$$

$$f(x^\star) = 0, \tag{8b}$$

$$0 \preceq \nu \perp -h(x^\star) \succeq 0. \tag{8c}$$

The first condition (Eq. 8a) is the stationarity of the Lagrangian for the primal variable, that is, $\nabla_x L(x,\lambda^\star,\nu^\star) = 0$. The second condition (Eq. 8b) is the stationarity of the Lagrangian for the first multipliers, that is, $\nabla_\lambda L(x^\star,\lambda,\nu^\star) = 0$; also know as *primal (equality) feasibility* condition. The third condition (Eq. 8b) is more involving: it combines *dual feasibility* $\nu \succeq 0$, *primal (inequality) feasibility* $h(x) \preceq 0$, and *complementary slackness* $\nu_i h_i = 0$ $(i = 1,\ldots,N_h)$.

There is a nice physical interpretation to the KKT conditions. Think about particles in the problem domain $\mathbb{R}^{N_x}$. The gradient $\nabla f_0(\cdot)$ can be interpreted as a force pushing those particles towards the point of minimum energy (the optimal point). However, we also have the forces $\nabla f(\cdot)$ and $\nabla h(\cdot)$ pushing those particles towards the feasible region. The optimal solution is a point where all these forces *balance* each other.

2.2 Equality-constrained problems

An equality-constrained convex program can be solved by applying root-finding algorithms to the KKT conditions,

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax - b = 0 \end{array} \quad \iff \quad \text{Find } (x^\star, \lambda^\star) \in \mathbb{R}^{N_x + N_f} \text{ such that } \begin{cases} \nabla f_0(x^\star) + A^\mathsf{T} \nu^\star = 0 \\ A^\mathsf{T} x^\star - b = 0 \end{cases} \tag{9}$$

That's it! We can now numerically solve an equality-constrained optimization problem.

There are some classes of problems for which the KKT conditions have a special *structure*, allowing us either to improve existing algorithms or to solve the optimization directly. A special case is equality-constrained QPs (when $f_0(x) = (1/2)x^\mathsf{T} P x + q^\mathsf{T} x + r$), which are equivalent to the root-finding problem

$$\text{Find } (x^\star, \lambda^\star) \in \mathbb{R}^{N_x + N_f} \text{ such that } \begin{bmatrix} P & A^\mathsf{T} \\ A^\mathsf{T} & \end{bmatrix} \begin{bmatrix} x^\star \\ \lambda^\star \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}. \tag{10}$$

Thus, solving equality-constrained QPs is equivalent to solving a linear system, which can be done using any sparsity-exploiting factor-solve method (e.g., using Cholesky or QR decompositions). Moreover, note that the coefficient matrix (known as the *KKT matrix*) is square, and it is nonsingular when $P \succ 0$. This system of $N_x + N_f$ equations is also known as the *KKT system*, and it is a building block of many modern solvers (including for nonconvex problems).

2.3 Inequality-constrained problems

Unfortunately, solving inequality-constrained problems is not as straightforward as solving equality-constrained problems. Specifically, the third KKT condition (Eq. 8c) prevents us from directly applying root-finding algorithms. However, we can still solve those problems using some smart tricks.

Interior-point methods

Convex programs can be approximated by the equality-constrained program

$$\text{minimize} \quad f_0(x) + \tfrac{1}{t}\phi(x) \tag{11a}$$

$$\text{subject to} \quad Ax + b = 0 \tag{11b}$$

given the log-*barrier function* $\phi(x) = -\sum_{i=1}^{N_f} \log(-h_i(x))$. This is a smooth approximation to an indicator function: it is equal to $-\infty$ when any inequality constraint is violated, and small when they are satisfied (Figure 2). Thus, the log-barrier enforces feasibility of the solutions at the expense of suboptimality (it can be understood as a *force* pushing particles away from the boundaries of the feasible set). Since the problem is now equality-constrained, it can be readily solved via root-finding algorithms (e.g., using Newton's method).

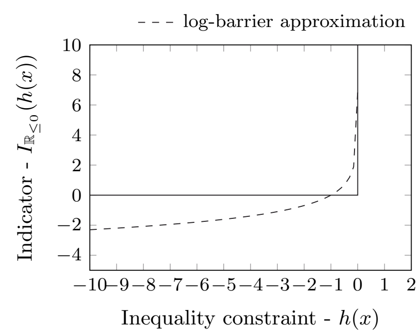

Figure 2. The log-barrier approximation.

Example 2.2 log-***barrier formulation of a constrained QP***

A convex QP can be approximated by the equality-constrained QP

$$\begin{array}{ll} \text{minimize} & (1/2)x^\mathsf{T} P x + q^\mathsf{T} x + r - \frac{1}{t}\sum_{i=1}^{N_x} \log(h_i - H_i^\mathsf{T} x) \\ \text{subject to} & Ax - b = 0. \end{array} \tag{12}$$

This approximated problem can be solved via the root-finding problem

$$\text{Find } x^\star \in \mathbb{R}^{N_x} \text{ such that } \begin{bmatrix} P & A^\mathsf{T} \\ A^\mathsf{T} & \end{bmatrix} \begin{bmatrix} x^\star \\ \lambda^\star \end{bmatrix} = \begin{bmatrix} \frac{1}{t} H^\mathsf{T} d(x^\star) - q \\ b \end{bmatrix}, \tag{13}$$

given the vector-valued function $d(x) = \big(1/(h_1 - H_1^\mathsf{T} x), \ldots, 1/(h_{N_h} - H_{N_h}^\mathsf{T} x)\big)$. Note that Problem (13) has a nonlinear system of equations (must be solved numerically, for instance using Newton's method).

A solution to the approximated Problem (11) can be shown to be (N_h/t)-suboptimal. Using the scaling factor $t > 0$, we can approximate the original problem with arbitrary precision (as $\frac{1}{t}\phi(x) \to I_{\mathbb{R}_{\leq 0}}(x)$ as $t \to \infty$). Roughly, the state-of-the-art approach to solve convex programs corresponds to solving a sequence of these approximated problems, increasing $t > 0$ each time until a desired accuracy is obtained: these methods are called *interior-point methods*.

References

[1] S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004. ISBN: 978-0-52-183378-3.