Online Lecture Notes

Prof. Boris Houska

May 17, 2022

1 Announcements

We have still three appointments left:

1. Today we will complete the lecture on Gauss-Newton Methods (Lecture 9) and cover the material on Equality Constrained Optimization (Lecture 10).
2. On this Thursday, we will complete Lecture 10 and have a Q&A + review of Lectures 7, 8, 9, and 10.
3. The final exam will be on Tuesday, May 24. This will be a 24h take home exam (as last time). The exam will be very similar to previous years. Material: the focus will be on
   (a) Lecture 7: Linear Equations Systems (Condition Numbers, LR, QR, and Cholesky Factorization)
   (b) Lecture 8: Newton Method, Convergence Proof, Globalization via Line Search
   (c) Lecture 9: Gauss-Newton for Least-Squares Regression, Convergence Analysis
   (d) Lecture 10: Newton-type methods for equality constrained optimization.
   (e) Homeworks 6 and 7.
   (f) There will be an online exercise session before the exam on Monday, May 23. Ask our TAs for details. Please attend!
4. This course runs for 12 weeks (= 3 credits) + 1 project credit (algorithmic differentiation homework)
5. Grading: 30% Homework, 30% Mid-Term, 40% Final Exam. We will send you the results of the exams in the coming weeks.

2 Equality Constrained Optimization

Today's lecture is about equality constrained optimization problems of the form

$$\min_x \; F(x) \quad \text{s.t.} \quad G(x) = 0$$

for twice Lipschitz-continuously differentiable functions $F : \mathbb{R}^n \to \mathbb{R}$ and $G : \mathbb{R}^n \to \mathbb{R}^m$, where we would have usually have $m \leq n$, since, otherwise, there are not enough variables to satisfy all equality constraints.

2.1 Example

Let us consider the case $n = 2$, $m = 1$, $F(x) = x_1^2 + x_2^2$, and $G(x) = x_1 + x_2 - 1$. There two ways to solve this problem explicitly. The first option is to eliminate one of the variables. For example, we could eliminate x_1 and find

$$x_1 + x_2 = 1 \qquad \Longrightarrow \qquad x_1 = 1 - x_2 \; .$$

In this case, we can substitute the parametric equation for x_1 into our objective,

$$\min_{x_2} \; (1 - x_2)^2 + x_2^2 ,$$

which leads to an unconstrained optimization problem. In this case, we can solve the stationarity condition

$$0 = -2(1 - x_2) + 2x_2 \qquad \Longrightarrow \qquad x_2 = \frac{1}{2} \; .$$

The result for x_1 can then be otained by substituting the result for x_2,

$$x_1 = 1 - x_2 = \frac{1}{2} \; .$$

In summary, the minimizer is at $\left(\frac{1}{2}, \frac{1}{2}\right)$.

The other option for arriving at the same result is to write the Lagrange optimality condition:

$$\begin{aligned} L(x,\lambda) &= F(x) + \lambda^\intercal G(x) = x_1^2 + x_2^2 + \lambda(x_1 + x_2 - 1) \\ 0 &= \nabla_x L(x,\lambda) = \begin{pmatrix} 2x_1 + \lambda \\ 2x_2 + \lambda \end{pmatrix} \\ 0 &= x_1 + x_2 - 1 \end{aligned} \tag{1}$$

We can summarize this equation system in the form

$$\begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

This yields

$$\begin{pmatrix} x_1 \\ x_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -1 \end{pmatrix}$$

The multiplier $-\lambda$ can be interpreted as the price of the constraint.

2.2 Newton-Type Methods for Equality Constrained Optimization

The second path for solving the above example can be generalized to derive a generic method for equality constrained optimization. The main idea is to introduce the Lagrangian function

$$L(x, \lambda) = F(x) + \lambda^\intercal G(x)$$

and write the first order optimality conditions in the form

$$0 = R(z) = \left(\begin{array}{c} \nabla_x L(x, \lambda) \\ G(x) \end{array} \right) \qquad \text{with} \qquad z = \left(\begin{array}{c} x \\ \lambda \end{array} \right) \in \mathbb{R}^{n+m} \ .$$

If we want to solve this equation system with respect to x and λ, we can apply a Newton-type method to the function $R : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$. The corresponding step direction Δz is obtained as

$$\Delta z = -M(z_k)^{-1} R(z) \ .$$

The final method is then updating $z_{k+1} = z_k + \alpha_k \Delta z_k$ with line search parameter $\alpha_k \in (0, 1]$.

2.3 Methods for constructing M in the context of equality constrained optimization

In order to discuss how to choose M, we first work out the exact Newton method, which is obtained by setting $M(z_k) = \frac{\partial}{\partial z} R(z_k)$. Notice that the Jacobian of R is given by

$$\begin{aligned} \frac{\partial}{\partial z} R(z) &= \frac{\partial}{\partial z} \left(\begin{array}{c} \nabla_x L(x, \lambda) \\ G(x) \end{array} \right) \\ &= \left(\begin{array}{cc} \nabla_{xx}^2 L(x, \lambda) & \nabla_{x\lambda}^2 L(x, \lambda) \\ G'(x) & 0 \end{array} \right) \end{aligned} \tag{2}$$

where $G' = \frac{\partial}{\partial x} G$ denotes the Jacobian of G. We can also work out the mixed second derivative term

$$\begin{aligned} \nabla_{x\lambda}^2 L(x, \lambda) &= \frac{\partial}{\partial \lambda} \nabla_x L(x, \lambda) \\ &= \frac{\partial}{\partial \lambda} \nabla_x \left[F(x) + \lambda^\intercal G(x) \right] \\ &= \frac{\partial}{\partial \lambda} \left[\nabla_x F(x) + G'(x)^\intercal \lambda \right] \\ &= G'(x)^\intercal \end{aligned} \tag{3}$$

Thus, in summary this means that

$$\frac{\partial}{\partial z} R(z) = \left(\begin{array}{cc} \nabla_{xx}^2 L(x, \lambda) & G'(x)^\intercal \\ G'(x) & 0 \end{array} \right) .$$

Notice that $\frac{\partial}{\partial z}R(z)$ is a symmetric matrix whose lower right block component is equal to 0. This particular structure can be exploited if we implement the exact Newton method. If we want to use a Newton type method instead, we can do this too by exploiting the structure of $\frac{\partial}{\partial z}R(z)$. A very popular choice for M has the form

$$\frac{\partial}{\partial z}R(z) \;=\; M(z) \;=\; \begin{pmatrix} H(z) & G'(x)^\intercal \\ G'(x) & 0 \end{pmatrix} ,$$

where G' is still denoting the exact Jacobian of G but $H \approx \nabla_{xx}^2 L(x,\lambda)$ denotes a positive definite Hessian approximation.

3 Summary

Notice that we don't go over all the slides of Lecture 10 today, but it will be enough to know all the material that are in this lecture notes. Some important naming conventions are:

- The function L is called the *Lagrange Function*
- The function R is called the *KKT Residuum*, where "KKT" stands for the three names Karush, Kuhn, and Tucker (they were the first introducing general optimality conditions for constrained optimization)
- The matrix
$$\frac{\partial}{\partial z}R(z) \;=\; \begin{pmatrix} \nabla_{xx}^2 L(x,\lambda) & G'(x)^\intercal \\ G'(x) & 0 \end{pmatrix} .$$
is called the *KKT matrix*. Notice that this matrix is invertible if $\nabla_{xx}^2 L(x,\lambda)$ is positve definite and $G'(x)$ has full rank.
- The condition that $G'(x^\star)$ has full rank at the optimal solution is known under the name "*Linear Independence Constraint Qualification*" (LICQ).
- The final Newton type method can be written in the form
$$z_{k+1} = z_k + \alpha_k \Delta z_k \qquad \text{with} \qquad \Delta z_k = M(z_k)^{-1} R(z_k) .$$
- In practice the matrix M is chosen to be of the form
$$M(z) \;=\; \begin{pmatrix} H(z) & G'(x)^\intercal \\ G'(x) & 0 \end{pmatrix}$$
where $H(z)$ is symmetric and positive definite and $G'(x)$ is the exact Jacobian.
- Notice that the convergence analysis of Newton type also applies here. We do get one step convergence if F is quadratic and G linear, since in this case R is linear. This means that equality constrained quadratic programming problems can solved by solving a linear equation system.

That's all that we need from Lecture 10 for the final exam!