Convex Optimization: Reading Notes 6

GKxx

May 29, 2022

1 Equality constrained minimization

We consider the minimization problems with only equality constraints, which are of the form

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & Ax = b. \end{aligned} \tag{1}$$

We may assume that f is convex and twice continuously differentiable. For the matrix $A \in \mathbb{R}^{p \times n}$ in the equality constraints, it is assumed that $\operatorname{rank} A = p < n$. This assumption is reasonable, in the sense that for $\operatorname{rank} A < p$ or $p \geqslant n$ some of the rows of A must be linearly expressible by others, which could be removed.

Suppose the optimal value p^* is finite and attained. The sufficient and necessary condition for x^* to be optimal is that

$$\begin{cases} x^* \in \operatorname{\mathbf{dom}} f, \quad Ax^* = b, \\ \nabla f(x^*) + A^{\mathsf{T}} v^* = 0 \quad \text{for some } v^*. \end{cases}$$

The equation $Ax^* = b$ indicates the *primal feasibility*, and $\nabla f(x^*) + A^{\mathsf{T}} v^* = 0$ indicates the *dual feasibility*.

Example 1.1 (Equality constrained convex quadratic minimization)**.** *The problem of the form in (1) with objective*

$$f(x) = \frac{1}{2} x^{\mathsf{T}} P x + q^{\mathsf{T}} x + r,$$

where $P \in \mathbb{S}_+^n$, is the convex quadratic minimization problem with equality constraints.

The optimality conditions for the convex quadratic problem could be written as

$$\begin{bmatrix} P & A^{\mathsf{T}} \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ v^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}.$$

The matrix

$$K = \begin{bmatrix} P & A^{\mathsf{T}} \\ A & 0 \end{bmatrix}$$

is called the *KKT matrix*. When the KKT matrix is nonsingular, there is a unique optimal primal-dual pair (x^*, v^*). In fact, the nonsingularity of K is equivalent to the following statements.

1. $\mathcal{N}(P) \cap \mathcal{N}(A) = 0$.

2. $Ax = 0, x \neq 0 \Rightarrow x^\mathsf{T} Px > 0$.

3. $F^\mathsf{T} P F \succ 0$, where $F \in \mathbb{R}^{n \times (n-p)}$ is a matrix for which $\mathcal{R}(F) = \mathcal{N}(A)$.

4. $P + A^\mathsf{T} Q A \succ 0$ for some $Q \succeq 0$.

Proof. Obviously, the first and the second statements are equivalent. The second and the third statements are equivalent in the sense that for all $Ax = 0, x \neq 0$, x must be of the form $x = Fy$ where $y \neq 0$, so that $x^\mathsf{T} Px = y^\mathsf{T} F^\mathsf{T} PFy > 0$.

Then we show the equivalence of the second and the fourth statements. When the second statement holds, pick $Q = I$ in the fourth statement and then for every x we have

$$x^\mathsf{T} \left(P + A^\mathsf{T} Q A \right) x = x^\mathsf{T} Px + (Ax)^\mathsf{T}(Ax) > 0,$$

so $P + A^\mathsf{T} Q A \succ 0$. Conversely, when $P + A^\mathsf{T} Q A \succ 0$ holds for some $Q \succeq 0$, it means that for every x we have

$$x^\mathsf{T} Px + x^\mathsf{T} A^\mathsf{T} Q A x > 0.$$

So for $Ax = 0, x \neq 0$ this gives $x^\mathsf{T} Px > 0$. Finally we show that these four statements are equivalent with the nonsingularity of K. For nonsingular K, the equation

$$K \begin{bmatrix} x \\ 0 \end{bmatrix} = \begin{bmatrix} Px \\ Ax \end{bmatrix} = 0$$

holds only when $x = 0$, which gives that $\mathcal{N}(P) \cap \mathcal{N}(A) = 0$. On the other hand, suppose K is singular, i.e. there exist x, z not both zero such that

$$\begin{bmatrix} P & A^\mathsf{T} \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} Px + A^\mathsf{T} z \\ Ax \end{bmatrix} = 0.$$

Multiplying $Px + A^\mathsf{T} z = 0$ on the left by x^T yields

$$x^\mathsf{T} Px + (Ax)^\mathsf{T} z = 0.$$

This reduces to $x^\mathsf{T} Px = 0$ when $Ax = 0$, and since $P \succeq 0$ it follows that $Px = 0$. This is possible only when $x = 0$, because $\mathcal{N}(P) \cap \mathcal{N}(A) = 0$, and then it leads to $A^\mathsf{T} z = 0$ which implies $z = 0$ for $\operatorname{rank} A = p$. This contradicts the assumption that x and z are not both zero. $\square$

It is also worth attention that when K is nonsingular, K must have n positive eigenvalues and p negative eigenvalues. The n positive eigenvalues are easy to see from the following interlacing theorem.

Theorem 1.2. *Let $A \in \mathbb{C}^{n \times n}$ be Hermitian partitioned as*

$$A = \begin{bmatrix} B & C \\ C^* & D \end{bmatrix}, \quad B \in \mathbb{C}^{m \times m}, D \in \mathbb{C}^{(n-m) \times (n-m)}, C \in \mathbb{C}^{m \times (n-m)}.$$

Then

$$\lambda_i(A) \leqslant \lambda_i(B) \leqslant \lambda_{i+n-m}(A)$$

for every $i = 1, \cdots, m$. Here $\lambda_i(X)$ denotes the i-th smallest eigenvalue of X.

1.3 Eliminating equality constraints

One general approach to solving equality constrained problems is to eliminate the equality constraints to reduce the problem to an unconstrained one. For the problem given in the form (1), one possible way of eliminating the equality constraints is to choose a matrix $F \in \mathbb{R}^{n \times (n-p)}$ and a vector $\hat{x} \in \mathbb{R}^n$ such that

$$\{x \mid Ax = b\} = \{Fz + \hat{x} \mid z \in \mathbb{R}^{n-p}\}.$$

We may choose $\hat{x}$ to be a solution of $Ax = b$, and F is any matrix satisfying

$$\mathcal{R}(F) = \mathcal{N}(A).$$

Note that the existence of such F is guaranteed in the convex quadratic problem when the KKT matrix is nonsingular. Then the reduced optimization problem is formed:

$$\min \quad \tilde{f}(z) = f(Fz + \hat{x}).$$

From its solution z^*, the solution of the original problem could be found by $x^* = Fz^* + \hat{x}$.

2 Newton's method

The general idea of finding an optimal solution of a problem is through a number of iterations, starting at point $x^{(0)}$, of the form

$$x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)},$$

with $f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right)$. In the iteration equation, $t^{(k)}$ is called the *step size* and $\Delta x^{(k)}$ is the *search direction*.

2.1 Interpretation: second-order approximation

The Newton's method gives a preferable way of iteration. There are a number of different interpretations of the Newton's step. One way is to derive it from optimizing a second-order approximation problem

$$\min_{v} \quad f(x) + \nabla f(x)^\mathsf{T} v + \frac{1}{2} v^\mathsf{T} \nabla^2 f(x) v$$
$$\text{s.t.} \quad A(x + v) = b.$$

Here the objective is given by the second-order Taylor's formula. Assuming the associated KKT matrix is nonsingular. The optimality conditions for this problem are

$$\begin{cases} \nabla f(x) + \nabla^2 f(x)^\mathsf{T} v + A^\mathsf{T} v = 0, \\ Ax + Av = b. \end{cases}$$

Noting that $Ax = b$, the above conditions could be rewritten as

$$\begin{bmatrix} \nabla^2 f(x) & A^\mathsf{T} \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ v \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}.$$

(Please discriminate between the 'ν' and 'v' here.) Hence the Newton step Δx_{nt} is characterized by

$$\begin{bmatrix} \nabla^2 f(x) & A^\mathsf{T} \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{nt} \\ \nu \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix},$$

where ν is the associated optimal dual variable for the quadratic problem.

2.2 Interpretation: linearized optimality conditions

The optimality conditions

$$\begin{cases} Ax^* = b \\ \nabla f\left(x^*\right) + A^\mathsf{T} v^* = 0 \end{cases}$$

could be approximated by substituting $x + \Delta x_{nt}$ for x^* and w for v^*, and then replacing the gradient term by its linearized approximation near x. These approximations yield

$$\begin{cases} A\left(x + \Delta x_{nt}\right) = b, \\ \nabla f\left(x + \Delta x_{nt}\right) + A^\mathsf{T} w \approx \nabla f(x) + \nabla^2 f(x)\Delta x_{nt} + A^\mathsf{T} w = 0. \end{cases}$$

Noting that $Ax = b$, we obtain

$$\begin{cases} A\Delta x_{nt} = 0, \\ \nabla^2 f(x)\Delta x_{nt} + A^\mathsf{T} w = -\nabla f(x). \end{cases}$$

2.3 The Newton decrement

The Newton decrement for the equality constrained problem is defined as

$$\lambda(x) = \left(\Delta x_{nt}^\mathsf{T} \nabla^2 f(x)\Delta x_{nt}\right)^{1/2}.$$

We can relate the Newton decrement to the quantity $f(x) - \inf_y \tilde{f}(y)$, where $\tilde{f}$ is the second-order approximation of f at x given by the Taylor's formula. We have that

$$f(x) - \inf_y \tilde{f}(y) = f(x) - \tilde{f}\left(x + \Delta x_{nt}\right) = \frac{1}{2}\lambda(x)^2.$$

Therefore, $\lambda(x)^2/2$ is an estimation of $f(x) - p^*$. Moreover, $\lambda(x)$ could also be seen as the norm of Δx_{nt}, in the quadratic norm defined by the Hessian

$$\|u\|_{\nabla^2 f(x)} = \left(u^\mathsf{T} \nabla^2 f(x)u\right)^{1/2}.$$

Moreover, Δx_{nt} is also a feasible descent direction