

1 Outline

In this lecture, we study

- the infeasible start Newton method,
- the primal-dual interior point method.

2 Infeasible start Newton method

Recall that Newton's method can be extended to solve the following equality constrained problem.

$$\begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && Ax = b. \end{aligned} \tag{24.1}$$

The update rule is that given a current iterate x_t , we obtain

$$x_{t+1} = x_t + d$$

where d is chosen to be an optimal solution to the following.

$$\begin{aligned} & \text{minimize} && f(x_t) + \nabla f(x_t)^\top d + \frac{1}{2} d^\top \nabla^2 f(x_t) d \\ & \text{subject to} && A(x_t + d) = b. \end{aligned} \tag{24.2}$$

Remember that based on the KKT conditions, we derived a necessary and sufficient condition for d as follows.

$$\begin{aligned} \nabla f(x_t) + \nabla^2 f(x_t) d + A^\top \mu &= 0, \\ A(x_t + d) &= b. \end{aligned}$$

Subject to $Ax_t = b$, this can be expressed as the following matrix system.

$$\begin{bmatrix} \nabla^2 f(x_t) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} d \\ \mu \end{bmatrix} = \begin{bmatrix} -\nabla f(x_t) \\ 0 \end{bmatrix}.$$

Here is our next question. What if the current iterate x_t is not feasible, meaning $Ax_t \neq b$? In this case, the corresponding matrix system for characterizing d is

$$\begin{bmatrix} \nabla^2 f(x_t) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} d \\ \mu \end{bmatrix} = - \begin{bmatrix} \nabla f(x_t) \\ Ax_t - b \end{bmatrix}. \tag{24.3}$$

Here, if the KKT matrix is invertible, we can deduce the desired direction d and obtain a new iterate $x_{t+1} = x_t + d$. This suggests that we may deal with infeasible iterates that are generated in intermediate steps. Then this raises one further question. Can we allow a sequence of infeasible

iterates? When x_t is infeasible and d is the associated direction with $x_t + d$ is feasible, instead of taking $x_t + d$, let us take

$$x_{t+1} = x_t + \eta d$$

for some step size η . Here, if $Ad \neq 0$ and $\eta \neq 1$, then $x_t + \eta d$ is not feasible. Nevertheless, as mentioned before, we can proceed the algorithm regardless of the feasibility of x_{t+1} .

For the remainder of this section, we use notation Δx to replace d to emphasize that the direction is the change we make. Moreover, note that we obtain a new dual variable μ every time we solve (24.3). Here, one may record the difference between the current dual variable and the new dual variable. Let us denote by $\Delta\mu$ the incremental change in the dual variable. Then the KKT conditions can be rewritten as

$$\begin{aligned}\nabla f(x) + \nabla^2 f(x) \Delta x + A^\top(\mu + \Delta\mu) &= 0, \\ A(x + \Delta x) &= b.\end{aligned}$$

Here, Δx and $\Delta\mu$ can be found by solving

$$\begin{bmatrix} \nabla^2 f(x) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\mu \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + A^\top \mu \\ Ax - b \end{bmatrix}. \quad (24.4)$$

This point of view suggests a primal-dual algorithm, Given a point x , our next point is given by

$$x + \eta \Delta x.$$

Following the update rule for the x variables, we may update the dual variable as

$$\mu + \eta \Delta\mu$$

where μ is the current dual variable. This is called a primal-dual method because we update both the original variable x and the dual variable μ at each iteration.

We stop the algorithm when Δx and $\Delta\mu$ become sufficiently small. This is equivalent to have

$$r(x, \mu) = \begin{bmatrix} \nabla f(x) + A^\top \mu \\ Ax - b \end{bmatrix} = - \begin{bmatrix} \nabla^2 f(x) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\mu \end{bmatrix}$$

sufficiently small.

Algorithm 1 Infeasible start Newton method

Initialize $t = 1$, x_1 , μ_1 , an accuracy level ϵ , and parameters $0 < \alpha < 1/2$ and $0 < \beta < 1$.

while $Ax_t \neq b$ or $\|r(x_t, \mu_t)\|_2 > \epsilon$ **do**

 Obtain Δx_t and $\Delta\mu_t$.

 Apply backtracking line search on $\|r\|_2$ with parameters α and β as follows.

 Set $k = 1$.

while $\|r(x_t + k\Delta x_t, \mu_t + k\Delta\mu_t)\|_2 > (1 - k\alpha)\|r(x_t, \mu_t)\|_2$ **do**

$k \leftarrow \beta k$.

end while

 Update $x_{t+1} = x_t + k\Delta x_t$ and $\mu_{t+1} = \mu_t + k\Delta\mu_t$.

$t \leftarrow t + 1$.

end while

Here, $r(x + \Delta x, \mu + \Delta \mu)$ can be expressed as

$$r(x + \Delta x, \mu + \Delta \mu) = \begin{bmatrix} \nabla f(x + \Delta x) + A^\top(\mu + \Delta \mu) \\ A(x + \Delta x) - b \end{bmatrix} \approx r(x, \mu) + \begin{bmatrix} \nabla^2 f(x) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \mu \end{bmatrix}.$$

Hence, computing Δx and $\Delta \mu$ can be interpreted as trying to make $r(x + \Delta x, \mu + \Delta \mu) \approx 0$.

3 Primal-dual interior point method

In the last lecture, we learned the barrier method for solving the following constrained convex minimization problem.

$$\begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && g_i(x) \leq 0, \quad i = 1, \dots, m, \\ & && Ax = b. \end{aligned} \tag{24.5}$$

For the barrier method, we used the log-barrier function given by

$$\psi(x) = - \sum_{i=1}^m \log(-g_i(x)).$$

Then by solving

$$\begin{aligned} & \text{minimize} && f(x) + \frac{1}{t} \psi(x) \\ & \text{subject to} && Ax = b \end{aligned} \tag{24.6}$$

for $t > 0$, we compute an optimal solution $x^*(t)$ for each t and construct the central path $\{x^*(t) : t > 0\}$. Moreover, we deduced the associated dual variables $\lambda_i^*(t)$ and $\mu^*(t)$ defined as

$$\lambda_i^*(t) = -\frac{1}{t \cdot g_i(x^*(t))}, \quad i = 1, \dots, m, \quad \mu^*(t) = \frac{\mu^*}{t}.$$

We also saw that $(x, \lambda, \mu) = (x^*(t), \lambda^*(t), \mu^*(t))$ satisfies the perturbed KKT conditions given by

$$\begin{aligned} \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) + A^\top \mu &= 0, \\ \lambda_i g_i(x) &= -\frac{1}{t}, \quad i = 1, \dots, m, \\ g_i(x) &\leq 0, \quad i = 1, \dots, m \\ Ax &= b, \\ \lambda_i &\geq 0, \quad i = 1, \dots, m. \end{aligned} \tag{24.7}$$

In this section, we will develop primal-dual methods for solving (24.5) based on the perturbed KKT conditions given in (24.7).

3.1 Primal-dual interpretation for the barrier method

For the barrier method, we have the condition that

$$\lambda_i = -\frac{1}{t g_i(x)}, \quad i = 1, \dots, m.$$

Plugging in this to (24.7), we deduce

$$\begin{aligned}\nabla f(x) - \frac{1}{t} \sum_{i=1}^m \frac{1}{g_i(x)} \nabla g_i(x) + A^\top \mu &= 0, \\ Ax - b &= 0.\end{aligned}$$

The barrier method solves (24.6) with the objective function.

$$h(x) = f(x) - \frac{1}{t} \sum_{i=1}^m \log(-g_i(x)).$$

Note that the perturbed KKT conditions are nothing but

$$\begin{aligned}\nabla h(x) + A^\top \mu &= 0, \\ Ax - b &= 0\end{aligned}$$

because

$$\nabla h(x) = \nabla f(x) - \frac{1}{t} \sum_{i=1}^m \frac{1}{g_i(x)} \nabla g_i(x).$$

In fact, this system comes up for the infeasible start Newton method. When we apply the infeasible start Newton method to solve (24.6), we proceed with the system

$$\begin{bmatrix} \nabla^2 h(x) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \mu \end{bmatrix} = - \begin{bmatrix} \nabla h(x) + A^\top \mu \\ Ax - b \end{bmatrix}.$$

Here, $\nabla^2 h(x)$ is given by

$$\nabla^2 h(x) = \nabla^2 f(x) + \frac{1}{t} \sum_{i=1}^m \frac{1}{g_i(x)^2} \nabla g_i(x) \nabla g_i(x)^\top - \frac{1}{t} \sum_{i=1}^m \frac{1}{g_i(x)} \nabla^2 g_i(x).$$

Basically, the perturbed KKT conditions characterize (x, μ) satisfying

$$r(x, \mu) = \begin{bmatrix} \nabla h(x) + A^\top \mu \\ Ax - b \end{bmatrix} = 0.$$

The infeasible start Newton method tries to find a pair (x, μ) with $r(x, \mu)$. As we mentioned, before the infeasible start Newton method is a primal-dual algorithm. Therefore, the barrier method with the infeasible start Newton method can be interpreted as a primal-dual method.

3.2 Primal-dual method by the perturbed KKT conditions

In the previous subsection, we observed that the perturbed KKT conditions with λ removed based on the barrier method leads to a primal-dual algorithm. In fact, we may design another primal-dual algorithm based on the perturbed KKT conditions (24.7) without removing the λ variables.

Let us use notations $g(x)$ and $Dg(x)$ to denote

$$g(x) = \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix}, \quad Dg(x) = \begin{bmatrix} \nabla g_1(x)^\top \\ \vdots \\ \nabla g_m(x)^\top \end{bmatrix}.$$

Then the equality conditions in (24.7) can be written as

$$\begin{aligned} \nabla f(x) + Dg(x)^\top \lambda + A^\top \mu &= 0, \\ -\text{Diag}(\lambda)g(x) - \frac{1}{t}\mathbf{1} &= 0, \\ Ax - b &= 0 \end{aligned} \tag{24.8}$$

where $\mathbf{1}$ denotes the vector of all ones and $\text{Diag}(v)$ denotes the diagonal matrix whose diagonal entries are given by the components of vector v . Then we define $r(x, \lambda, \mu)$ as

$$r(x, \lambda, \mu) = \begin{bmatrix} r_{\text{dual}}(x, \lambda, \mu) \\ r_{\text{central}}(x, \lambda, \mu) \\ r_{\text{primal}}(x, \lambda, \mu) \end{bmatrix} = \begin{bmatrix} \nabla f(x) + Dg(x)^\top \lambda + A^\top \mu \\ -\text{Diag}(\lambda)g(x) - \frac{1}{t}\mathbf{1} \\ Ax - b \end{bmatrix}.$$

A primal-dual method would seek to update (x, λ, μ) as follows.

$$x \rightarrow x + \eta \Delta x, \quad \lambda \rightarrow \lambda + \eta \Delta \lambda, \quad \mu \rightarrow \mu + \eta \Delta \mu.$$

How do we find the increments Δx , $\Delta \lambda$, and $\Delta \mu$? Note that

$$\begin{aligned} & r(x + \Delta x, \lambda + \Delta \lambda, \mu + \Delta \mu) \\ &= \begin{bmatrix} \nabla f(x + \Delta x) + Dg(x + \Delta x)^\top (\lambda + \Delta \lambda) + A^\top (\mu + \Delta \mu) \\ -\text{Diag}(\lambda + \Delta \lambda)g(x + \Delta x) - \frac{1}{t}\mathbf{1} \\ A(x + \Delta x) - b \end{bmatrix} \\ &\approx r(x, \lambda, \mu) + \begin{bmatrix} \nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 g_i(x) & Dg(x)^\top & A^\top \\ -\text{Diag}(\lambda)Dg(x) & -\text{Diag}(g(x)) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta \mu \end{bmatrix} \end{aligned}$$

Basically, we find $(\Delta x, \Delta \lambda, \Delta \mu)$ so that $r(x + \Delta x, \lambda + \Delta \lambda, \mu + \Delta \mu) \approx 0$. We may achieve this by solving

$$\begin{bmatrix} \nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 g_i(x) & Dg(x)^\top & A^\top \\ -\text{Diag}(\lambda)Dg(x) & -\text{Diag}(g(x)) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta \mu \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + Dg(x)^\top \lambda + A^\top \mu \\ -\text{Diag}(\lambda)g(x) - \frac{1}{t}\mathbf{1} \\ Ax - b \end{bmatrix}.$$

Based on this, we may deduce a primal-dual algorithm. Given (x_t, λ_t, μ_t) , we obtain

$$(x_{t+1}, \lambda_{t+1}, \mu_{t+1}) = (x_t + \eta \Delta x_t, \lambda_t + \eta \Delta \lambda_t, \mu_t + \eta \Delta \mu_t)$$

for some step size $\eta > 0$.

3.3 Primal-dual interior point method

How do we guarantee convergence of the primal-dual algorithm? Suppose that

$$\begin{aligned} \nabla f(x) + Dg(x)^\top \lambda + A^\top \mu &= 0, \\ Ax - b &= 0 \end{aligned} \tag{24.9}$$

and that $g_i(x) \leq 0$ and $\lambda_i \geq 0$ for $i = 1, \dots, m$. This means that x is feasible to (24.5) and that

$$L(x, \lambda, \mu) = \min_x L(x, \lambda, \mu) = q(\lambda, \mu).$$

Here, we have

$$f(x) - q(\lambda, \mu) = - \sum_{i=1}^m \lambda_i g_i(x) - \mu^\top (Ax - b) = - \sum_{i=1}^m \lambda_i g_i(x).$$

As the Lagrangian duality implies that

$$f(x) - \min \{f(x) : g_i(x) \leq 0, i = 1, \dots, m, Ax = b\} \leq f(x) - q(\lambda, \mu),$$

we know that

$$- \sum_{i=1}^m \lambda_i g_i(x)$$

provides an optimality gap. However, the infeasible start Newton method does not guarantee feasibility for intermediate iterations. Therefore, $-\sum_{i=1}^m \lambda_i g_i(x)$ is not necessarily an upper bound on the optimality gap if (24.9) is not satisfied. Nevertheless, the term $-\sum_{i=1}^m \lambda_i g_i(x)$ provides a proxy for the optimality gap. Based on this observation, we may deduce the following algorithm.

Algorithm 2 Primal-dual interior point method

Initialize x_1 with $g_i(x_1) < 0$ for $i = 1, \dots, m$, $\lambda^1 > 0$, $\alpha > 1$, and an accuracy level ϵ .

Set $\delta_1 = -\sum_{i=1}^m \lambda_i^1 (g_i(x_1))$.

while $\delta_k > \epsilon$ or $(\|r_{\text{primal}}(x_k, \lambda^k, \mu_k)\|_2^2 + \|r_{\text{dual}}(x_k, \lambda^k, \mu_k)\|_2^2)^{1/2} > \epsilon$ **do**

 Set $t = \alpha m / \delta_k$

 Obtain Δx_k , $\Delta \lambda^k$, and $\Delta \mu_k$.

 Apply backtracking line search to determine step size η_k .

 Update $(x_{k+1}, \lambda^{k+1}, \mu_{k+1}) = (x_k + \eta \Delta x_k, \lambda^k + \eta \Delta \lambda^k, \mu_k + \eta \Delta \mu_k)$.

 Set $\delta^{k+1} = -\sum_{i=1}^m \lambda_i^{k+1} g_i(x_{k+1})$.

$k \leftarrow k + 1$.

end while

Here, the backtracking line search needs to find η such that $g_i(x_{k+1}) < 0$ and $\lambda_i^k > 0$ for $i = 1, \dots, m$.