Topic 6 Equality-constrained minimization.

$$\min f(x)$$
$$\text{s.t. } Ax = b$$

$f: \mathbb{R}^n \to \mathbb{R}$ twice different convex

$A \in \mathbb{R}^{p \times n}$ $(p < n)$

$\operatorname{rank} A = p$ (under determined)

Assume an x^* exists, and $p^* = f(x^*)$. $(p^* > -\infty)$.

+) Recall the optimality conditions:

x^* is optimal iff $\exists\, \nu^*$ such that $(\nu^* \in \mathbb{R}^p)$

$$\nabla f(x^*) + A^T \nu^* = 0 \quad \text{(dual feasible)}$$
$$Ax^* = b. \quad \text{(primal feasible)}$$

The KKT conditions give $n+p$ equations in $n+p$ variab $(x^*$ and $\nu^*)$.

+) To solve an equality-constrained problem, we can either
- Eliminate the equality constraint and solve the resulting unconstrained problem

or
- Solve the dual problem (assuming the dual function is twice differentiable)

Often keeping the constraints is preferred as it can give so structure to the problem that helps simplify computation.

+) Eliminating equality constraints:

Given $A \in \mathbb{R}^{p \times n}$, find a matrix $F \in \mathbb{R}^{n \times (n-p)}$ in the null space of A, and a feasible vector $\hat{x} \in \mathbb{R}^n$ such that

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

Here $A\hat{x} = b$ and $AF = 0$ $(\mathcal{R}(F) = \mathcal{N}(A))$.

Then form $\min \tilde{f}(z) = f(Fz + \hat{x})$

This is an unconstrained problem, $z \in \mathbb{R}^{n-p}$.

- From solution z^*, we obtain x^* and ν^* as

$$x^* = Fz^* + \hat{x}$$
$$\nu^* = -(AA^T)^{-1} A \nabla f(x^*)$$

This is because

$$\begin{bmatrix} F^T \\ A \end{bmatrix} \left(\nabla f(x^*) - A^T (AA^T)^{-1} A \nabla f(x^*) \right) = 0$$

where $F^T \nabla f(x^*) = \nabla \tilde{f}(z^*) = 0$. Since $\begin{bmatrix} F^T \\ A \end{bmatrix}$ is nonsingular, we must have

$$\nabla f(x^*) - A^T (AA^T)^{-1} A \nabla f(x^*) = 0 = \nabla f(x^*) + A^T \nu^*$$

- Note that if $T \in \mathbb{R}^{(n-p) \times (n-p)}$ is non singular then $\tilde{F} = FT$ is also a suitable elimination matrix. Hence the choice of F is not unique.

+) <u>Solving the problem via the dual:</u>

$$\mathcal{L}(x, \nu) = f(x) + \nu^T (Ax - b)$$

$$\rightarrow g(\nu) = -b^T \nu + \inf_x \left(f(x) + \nu^T A x \right)$$
$$= -b^T \nu - \sup_x \left(-(A^T \nu)^T x - f(x) \right)$$
$$= -b^T \nu - f^*(-A^T \nu) \quad \text{(conjugate function)}.$$

Thus the dual problem is

$$\max \; -b^T \nu - f^*(-A^T \nu)$$

Since strong duality holds (affine constraint with the optimal point being feasible) $\rightarrow g(\nu^*) = p^*$

Reconstruct x^* from ν^* may not be straightforward.

Example: $\min -\sum_{i=1}^{n} \log x_i$ (equality constrained analytic center)

s.t. $Ax = b$, $A \in \mathbb{R}^{p\times n}$.

Here implicitly $x \succ 0$.

Using the conjugate function

$$f^*(y) = \sum_{i=1}^{n} (-1 - \log(y_i)) = -n - \sum_{i=1}^{n} \log(-y_i)$$

the dual problem is

$$\max \quad -b^T\nu + n + \sum_{i=1}^{n} \log(A^T\nu)_i$$

with implicit constraint $A^T\nu \succ 0$.

The optimal primal and dual values are related by solving the dual feasibility equation:

$$\nabla f(x) + A^T\nu = 0 \Leftrightarrow -\frac{1}{x_i} + (A^T\nu)_i = 0 \quad \forall i = 1 \ldots$$

$$\to \quad x_i(\nu) = \frac{1}{(A^T\nu)_i} \quad \forall\, i = 1 \ldots n.$$

+) Equality constrained quadratic form: (special & important case)

$$\min \frac{1}{2} x^T P x + q^T x + r \qquad P \in S_+^n$$

$$\text{s.t.} \quad Ax = b \qquad A^T \in \mathbb{R}^{n\times p}$$

The optimality conditions are

$$Ax^* = b, \quad Px^* + q + A^T\nu^* = 0$$

which is equivalent to

$$\underbrace{\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}}_{\text{KKT matrix}} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

If the KKT matrix is nonsingular, there is a unique optimal primal-dual pair (x^*, ν^*). (Holds if $P \succ 0$).

If the KKT matrix is singular but the KKT system is solvable, any (x^*, ν^*) satisfies the KKT system will be optimal.

If the KKT system is not solvable, the problem is unbounded below or is infeasible.

+) <u>Newton's method with equality constraints:</u>

This is almost the same as Newton's method without constraints, except that all points $x^{(k)}$ must be feasible. Specifically:

(i) $x^{(0)}$ must be feasible: $Ax^{(0)} = b$

(ii) Newton step must be in a feasible direction:

$$A \Delta x_{nt} = 0.$$

- Second-order approximation:

$$\min f(x)$$
$$\text{s.t. } Ax = b$$

Replacing the objective by its second-order Taylor approx.:

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

Then the Newton step Δx_{nt} and the dual variable associated with the quadratic problem must satisfy

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

- Again $(\Delta x_{nt}, w)$ can be viewed as the solution to linearized optimality conditions.

$$\begin{cases} Ax^* = b \\ \nabla f(x^*) + A^T \nu^* = 0 \end{cases} \longrightarrow \begin{cases} A(x+v) = 0 \\ \nabla f(x) + \nabla^2 f(x) v + A^T w \cong 0 \end{cases}$$

∘ The Newton decrement

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

which is similar to the unconstrained case.
The meaning and use are the same as the unconstrained ca

Note with equality constraints, in general,

$$\lambda(x)^2 \neq \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)$$

∘ Feasibility descent direction:
Suppose $Ax = 0$. Then $v \in \mathbb{R}^n$ is a feasible direction if

$$Av = 0.$$

∘ Affine invariance: We can also show that the Newton step is affine invariant even with equality constraints:

If $x = Ty \rightarrow \Delta x_{nt} = T \Delta y_{nt}$
($T \in \mathbb{R}^{n \times n}$, T nonsingular).

∘ Feasible Newton's method:

given a starting point $x \in \operatorname{dom} f$ with $Ax = b$, and $\varepsilon >$
repeat
1. Compute the Newton step and decrement $\lambda(x)$
   solving $\begin{bmatrix} \nabla^2 f(x) & A^T \\ A^T & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{nt} \\ \omega \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}$.
2. Stopping criterion: quit if $\frac{\lambda^2}{2} \leq \varepsilon$.
3. Line search: choose t via backtracking
4. Update: $x^+ = x + t \Delta x_{nt}$.

Convergence analysis and property are similar to the unconstrained case.

+) Newton's method and elimination:

For the reduced problem (after eliminating equality constraints):

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

where $A\hat{x} = b$, $AF = 0$ $(\mathcal{R}(F) = \mathcal{N}(A))$. $F \in \mathbb{R}^{n\times(n-p)}$
variable is $z \in \mathbb{R}^{n-p}$.

Newton's method for $\tilde{f}(z)$ starts at $z^{(0)}$, generates $z^{(k)}$.

Newton's method with equality constraints produces:

$$x^{(k+1)} = Fz^{(k)} + \hat{x}$$

That is, $\Delta x_{nt} = F \Delta z_{nt}$.

+) Infeasible start Newton method:

- This is a generalization of Newton's method that works with initial points and iterates that are infeasible

Recall the optimality conditions:

$$Ax^* = b \quad , \quad \nabla f(x^*) + A^T \nu^* = 0$$

- Denote x as the current point, which may not be feasible.

We want to find Δx so that $x + \Delta x$ approximately satisfies the optimality conditions.

Substitute $x + \Delta x \to x^*$
$\qquad\qquad\;\; w \to \nu^*$

and use first order Taylor approximation:

$$\nabla f(x + \Delta x) \approx \nabla f(x) + \nabla^2 f(x) . \Delta x$$

We obtain:

$$\begin{cases} A(x+\Delta x) = b \\ \nabla f(x) + \nabla^2 f(x)\Delta x + A^T w = 0 \end{cases}$$

Rewrite in matrix form:

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ -Ax+b \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ Ax-b \end{bmatrix}$$

This equation is similar to the feasible method but with $Ax-b$ in place of 0.

$Ax-b$: residual vector for the feasibility constraint.

- Residual reduction and full step feasibility property:

Denote $r_p = Ax-b$ as the primal residual.

Note that the Newton step always satisfy

$$A(x+\Delta x_{nt}) = b.$$

- Thus if the step length $t=1$ is taken, then the following iterate will be feasible.
- Once an iterate is feasible, it will stay feasible for all the following iterates (see Homework P10.8)

- To see the effect on the residual during the damped phase (where $t<1$), note that the next iterate is

$$x^+ = x + t\Delta x_{nt}$$

→ residual at next iterate is

$$r_p^+ = A(x+t\Delta x_{nt}) - b = (1-t)(Ax-b) = (1-t)r_p.$$

Thus after $(k-1)$ iterations in the damped phase, we have

$$r_p^{(k)} = \left(\prod_{i=1}^{k-1}\left(1-t^{(i)}\right)\right) r^{(0)}.$$

Thus the residual scales down at each iterate. Once a full step is taken (eg. in the quadratic phase) then the residual is 0 and all future iterates are primal feasible.

+) Infeasible start Newton's method.

Define $r(x,\nu) = (r_d(x,\nu), r_p(x,\nu))$ (residual vector)

where $\begin{cases} r_d(x,\nu) = \nabla f(x) + A^T\nu & \text{dual residual} \\ r_p(x,\nu) = Ax - b & \text{primal residual} \end{cases}$

Given a starting point $x \in \text{dom} f$, ν, tolerance $\varepsilon > 0$, $\alpha \in (0, 1/2)$ and $\beta \in (0,1)$

repeat

1. Compute primal and dual steps $\Delta x_{nt}, \Delta\nu_{nt}$
2. Backtracking line search:
   $t := 1$
   while $\|r(x + t\Delta x_{nt}, \nu + t\Delta\nu_{nt})\|_2 > (1-\alpha t)\|r(x,\nu)\|_2$
   $t := \beta t$
3. Update $x := x + t\Delta x_{nt}$
   $\nu := \nu + t\Delta\nu_{nt}$

until $Ax = b$ and $\|r(x,\nu)\|_2 \leq \varepsilon$

A few notes on this infeasible start Newton method:

- It is not a descent method, we could have $f(x^{(k+1)}) > f(x^{(k)})$
- This is an instance of a primal-dual method in which we update both the primal x and dual ν at each iteration to approximately satisfy the optimality condition.
- The primal & dual updates need not be feasible at each step.
- Stopping criterion is based on the norm of the residual left, since this norm decreases in the Newton direction.

+) Solving the KKT conditions:

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = - \begin{bmatrix} g \\ h \end{bmatrix}$$

Elimination method (H nonsingular)

$$A H^{-1} A^T w = h - A H^{-1} g$$

$$H v = -(g + A^T w).$$

H singular: write

$$\begin{bmatrix} H + A^T Q A & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = - \begin{bmatrix} g + A^T Q h \\ h \end{bmatrix}$$

for $Q \succeq 0$, $H + A^T Q A$ nonsingular.