Equality Constrained Minimization

Lecture 12, Nonlinear Programming

National Taiwan University

December 27, 2016

Table of contents

Algorithms for Convex Optimization Problems (1/2)

- Our goal: learn algorithms that solve convex optimization problems efficiently:

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, 2, ..., m \\ & h_i(x) = 0, \quad i = 1, 2, ..., p \end{aligned}$$

where $f_i(x), 0 \leq i \leq m$ are convex functions and $h_i(x), 1 \leq i \leq p$ are affine functions.

Algorithms for Convex Optimization Problems (2/2)

- We have learned several descent methods for unconstrained convex optimization problems.
  - In particular, the Newton's method has the best convergence properties among them.
- In this lecture, we will study methods for solving a convex optimization problem with equality constraints, including an extension of the Newton's method.
- In the next lecture, we will study interior-point methods that solve convex optimization problems with both equality and inequality constraints, using the methods described here.

Equality Constrained Minimization Problems (1/2)

- We will describe methods for solving a convex optimization problem with equality constraints,

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b, \end{array}$$

  where $f : \mathbf{R}^n \to \mathbf{R}$ is convex and twice continuously differentiable, and $A \in \mathbf{R}^{p \times n}$ with $\operatorname{rank} A = p < n$.
- We assume that an optimal solution x^* exists, and use p^* to denote the optimal value, $p^* = \inf \{f(x) \mid Ax = b\} = f(x^*)$.

Equality Constrained Minimization Problems (2/2)

- Possible approaches:
  1. Reformulate the equality constrained problem into an equivalent unconstrained problem (by eliminating the equality constraints).
  2. Solve the dual problem using an unconstrained minimization method (if the dual function is twice differentiable), and then recover the solution of the equality constrained problem from the dual solution.
  3. Solve the KKT optimality conditions of the equality constrained problem.

Eliminating equality constraints (1/2)

- One general approach to solving the equality constrained problem is to eliminate the equality constraints, and then solve the resulting unconstrained problem using methods for unconstrained minimization.
- We first find a matrix $F \in \mathbf{R}^{n\times(n-p)}$ and vector $\hat{x} \in \mathbf{R}^n$ that parametrize the (affine) feasible set:

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

Eliminating equality constraints (2/2)

- Here $\hat{x}$ can be chosen as any particular solution of $Ax = b$, and $F \in \mathbf{R}^{n \times (n-p)}$ is any matrix whose range is the nullspace of A. We then form the reduced or eliminated optimization problem

$$\text{minimize} \quad \tilde{f}(z) = f(Fz + \hat{x}),$$

  which is an unconstrained problem with variable $z \in \mathbf{R}^{n-p}$.
- From its solution z^*, we can find the solution of the equality constrained problem as $x^* = Fz^* + \hat{x}$.

Example – Optimal allocation with resource constraint (1/2)

- We consider the problem

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{i=1}^{n} f_i(x_i) \\ \text{subject to} & \displaystyle\sum_{i=1}^{n} x_i = b, \end{array}$$

  where the functions $f_i : \mathbf{R} \to \mathbf{R}$ are convex and twice differentiable, and $b \in \mathbf{R}$ is a problem parameter.

- We interpret this as the problem of optimally allocating a single resource, with a fixed total amount b (the budget) to n otherwise independent activities.

Example – Optimal allocation with resource constraint (2/2)

- We can eliminate x_n using the parametrization $x_n = b - x_1 - \ldots - x_{n-1}$, which corresponds to the choices

$$\hat{x} = be_n, \quad F = \begin{bmatrix} I_{n-1} \\ -\mathbf{1}^T \end{bmatrix} \in \mathbf{R}^{n \times (n-1)}.$$

- The reduced problem is then

$$\text{minimize} \quad f_n(b - x_1 - \ldots - x_{n-1}) + \sum_{i=1}^{n-1} f_i(x_i),$$

with variables $x_1, \ldots, x_{n-1}$.

Choice of elimination matrix (1/2)

- There are many possible choices for the elimination matrix F, which can be chosen as any matrix in $\mathbf{R}^{n\times(n-p)}$ with $\mathcal{R}(F) = \mathcal{N}(A)$.
- If F is one such matrix, and $T \in \mathbf{R}^{(n-p)\times(n-p)}$ is nonsingular, then $\tilde{F} = FT$ is also a suitable elimination matrix, since

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

- Conversely, if F and $\tilde{F}$ are any two suitable elimination matrices, then there is some nonsingular T such that $\tilde{F} = FT$.

Choice of elimination matrix (2/2)

- If we eliminate the equality constraints using F, we solve the unconstrained problem

$$\text{minimize } f(Fz + \hat{x}),$$

while if $\tilde{F}$ is used, we solve the unconstrained problem

$$\text{minimize } f(\tilde{F}\tilde{z} + \hat{x}) = f(F(T\tilde{z}) + \hat{x}).$$

- This problem is equivalent to the one above, and is simply obtained by the change of coordinates $z = T\tilde{z}$.
- In other words, changing the elimination matrix can be thought of as changing variables in the reduced problem.

Solving equality constrained problems via the dual (1/2)

- Another approach to solving an equality-constrained problem is to solve the dual, and then recover the optimal primal variable x^*. The dual function of the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array}$$

can be written as

$$\begin{aligned} g(\nu) &= -b^T\nu + \inf_x (f(x) + \nu^T Ax) \\ &= -b^T\nu - \sup_x \left((-A^T\nu)^T x - f(x)\right) \\ &= -b^T\nu - f^*(-A^T\nu), \end{aligned}$$

Solving equality constrained problems via the dual (2/2)

where f^* is the conjugate of f, so the dual problem is

$$\text{maximize} \quad -b^T\nu - f^*(-A^T\nu).$$

- Since by assumption there is an optimal point, the problem is strictly feasible, so Slater's condition holds. Therefore strong duality holds, and the dual optimum is attained, i.e., there exists a ν^* with $g(\nu^*) = p^*$.
- If the dual function g is twice differentiable, then the methods for unconstrained minimization described in chapter 9 can be used to maximize g.
- Once we find an optimal dual variable ν^*, we may attempt to reconstruct an optimal primal solution x^* from it.

Example – Equality constrained analytic center (1/3)

- We consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) = -\sum_{i=1}^{n} \log x_i \\ \text{subject to} & Ax = b, \end{array}$$

where $A \in \mathbf{R}^{p \times n}$, with implicit constraint $x \succ 0$.

Example – Equality constrained analytic center (2/3)

- Using

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

(with **dom** $f^* = -\mathbf{R}^n_{++}$), the dual problem is

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

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

Example – Equality constrained analytic center (3/3)

- Here we can easily solve the dual feasibility equation, i.e., find the x that minimizes $L(x, \nu)$:

$$\nabla f(x) + A^T \nu = -(1/x_1, \ldots, 1/x_n) + A^T \nu = 0,$$

and so

$$x_i(\nu) = 1/(A^T \nu)_i.$$

Solving the KKT Conditions

- From KKT optimality conditions, a point $x^* \in \textbf{dom } f$ is optimal for the problem if and only if there is a $\nu^* \in \mathbf{R}^p$ such that
$$Ax^* = b, \quad \nabla f(x^*) + A^T \nu^* = 0,$$
which is a set of $n+p$ equations in the $n+p$ variables x^*, ν^*.
- The first set of equations, $Ax^* = b$, are called the **primal feasibility equations**, which are linear.
- The second set of equations, $\nabla f(x^*) + A^T \nu^* = 0$, are called the **dual feasibility equations**, and are in general nonlinear.
- We consider first the special case that these equations are also linear, that is, $\nabla f(x) = Px + q$ for some $P \in \mathbf{S}^n_+$ and $q \in \mathbf{R}^n$.

Equality constrained convex quadratic minimization (1/2)

- Consider the equality constrained convex quadratic minimization problem

$$\begin{array}{ll} \text{minimize} & f(x) = (1/2)x^T Px + q^T x + r \\ \text{subject to} & Ax = b, \end{array}$$

  where $P \in \mathbf{S}_+^n$ and $A \in \mathbf{R}^{p \times n}$.
- This problem is important on its own, and also because it forms the basis for an extension of Newton's method to equality constrained problems.

Equality constrained convex quadratic minimization (2/2)

- Here the optimality conditions are

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

which we can write as

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

- This set of $n+p$ linear equations in the $n+p$ variables x^*, ν^* is called the **KKT system** for the equality constrained quadratic optimization problem.
- The coefficient matrix is called the **KKT matrix**.

Singularity of the KKT matrix (1/2)

- When the KKT matrix

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

is nonsingular, there is a unique optimal primal-dual pair (x^*, ν^*).

- If the KKT matrix is singular, but the KKT system

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

is solvable, any solution yields an optimal pair (x^*, ν^*).

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

Singularity of the KKT matrix (2/2)

- In this case there exist $v \in \mathbf{R}^n$ and $w \in \mathbf{R}^p$ such that

$$Pv + A^T w = 0, \quad Av = 0, \quad -q^T v + b^T w > 0.$$

- Let $\hat{x}$ be any feasible point. Then, the point $x = \hat{x} + tv$ is feasible for all t and

$$\begin{aligned} f(\hat{x} + tv) &= f(\hat{x}) + t(v^T P\hat{x} + q^T v) + (1/2)t^2 v^T Pv \\ &= f(\hat{x}) + t(-\hat{x}^T A^T w + q^T v) - (1/2)t^2 w^T Av \\ &= f(\hat{x}) + t(-b^T w + q^T v), \end{aligned}$$

which decreases without bound as $t \to \infty$.

Conditions on Nonsingularity of the KKT matrix

- Recall our assumption that $P \in \mathbf{S}_+^n$ and rank $A = p < n$. There are several conditions equivalent to nonsingularity of the KKT matrix:
  - $\mathcal{N}(P) \bigcap \mathcal{N}(A) = 0$, i.e., P and A have no nontrivial common nullspace.
  - $Ax = 0,\ x \neq 0 \implies x^T P x > 0$, i.e., P is positive definite on the nullspace of A.
  - $F^T P F \succ 0$, where $F \in \mathbf{R}^{n \times (n-p)}$ is a matrix for which $\mathcal{R}(F) = \mathcal{N}(A)$.
- As an important special case, we note that if $P \succ 0$, the KKT matrix must be nonsingular.

Newton's method with equality constraints

- In this section we describe an extension of Newton's method to include equality constraints.
- The method is almost the same as Newton's method without constraints, except for two differences:
  1. The initial point must be feasible (i.e., satisfy $x \in \mathbf{dom}\ f$ and $Ax = b$).
  2. The definition of Newton step is modified to take the equality constraints into account.
- In particular, we make sure that the Newton step Δx_{nt} is a feasible direction, i.e., $A\Delta x_{\mathrm{nt}} = 0$.

Newton Step Defined via 2nd-Order Approximation (1/3)

- To derive the Newton step Δx_{nt} for the equality constrained problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b, \end{array}$$

at the feasible point x, we replace the objective with its second-order Taylor approximation near x, to form the problem

$$\begin{array}{ll} \text{minimize} & \hat{f}(x+v) = f(x) + \nabla f(x)^T v + (1/2) v^T \nabla^2 f(x) v \\ \text{subject to} & A(x+v) = b, \end{array}$$

with variable v, which is a (convex) quadratic minimization problem with equality constraints, and can be solved analytically.

Newton Step Defined via 2nd-Order Approximation (2/3)

- We define Δx_{nt}, the Newton step at x, as the solution of the convex quadratic problem

$$\begin{array}{ll} \text{minimize} & \hat{f}(x+v) = f(x) + \nabla f(x)^T v + (1/2) v^T \nabla^2 f(x) v \\ \text{subject to} & A(x+v) = b, \end{array}$$

  assuming the associated KKT matrix is nonsingular.

- Therefore, the Newton step Δx_{nt} is characterized by

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

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

Newton Step Defined via 2nd-Order Approximation (3/3)

- The Newton step Δx_{nt} is what must be added to x to solve the problem when the quadratic approximation is used in place of f. It is defined only at points for which the KKT matrix is nonsingular.
- As in Newton's method for unconstrained problems, we observe that when the objective f is exactly quadratic, the Newton update $x + \Delta x_{\mathrm{nt}}$ exactly solves the equality constrained minimization problem, and in this case the vector w is the optimal dual variable for the original problem.
- This suggests, as in the unconstrained case, that when f is nearly quadratic, $x + \Delta x_{\mathrm{nt}}$ should be a very good estimate of the solution x^*, and w should be a good estimate of the optimal dual variable ν^*.

The Newton decrement (1/3)

- We define the Newton decrement for the equality constrained problem as
$$\lambda(x) = (\Delta x_{\mathrm{nt}}^T \nabla^2 f(x) \Delta x_{\mathrm{nt}})^{1/2},$$
which is exactly the same expression as in the unconstrained case.
- The Newton decrement $\lambda(x)$ is the norm of the Newton step, in the norm determined by the Hessian $\nabla^2 f(x)$, i.e.,
$$\lambda(x) = ||\Delta x_{\mathrm{nt}}||_{\nabla^2 f(x)}.$$

The Newton decrement (2/3)

- Let

$$\hat{f}(x+v) = f(x) + \nabla f(x)^T v + (1/2) v^T \nabla^2 f(x) v$$

be the second-order Taylor approximation of f at x. The difference between $f(x)$ and the minimum of the second-order model satisfies

$$f(x) - \inf\left\{\hat{f}(x+v) \mid A(x+v) = b\right\} = \lambda(x)^2/2,$$

exactly as in the unconstrained case.

The Newton decrement (3/3)

- This means that, as in the unconstrained case, $\lambda(x)^2/2$ gives an estimate of $f(x) - p^*$, based on the quadratic model at x, and also that $\lambda(x)$ (or a multiple of $\lambda(x)^2$) serves as the basis of a good stopping criterion.
- The Newton decrement comes up in the line search as well, since the directional derivative of f in the direction Δx_{nt} is

$$\frac{d}{dt} f(x + t\Delta x_{\mathrm{nt}})\bigg|_{t=0} = \nabla f(x)^T \Delta x_{\mathrm{nt}} = -\lambda(x)^2,$$

as in the unconstrained case.

Feasible descent direction

- Suppose that $Ax = b$. We say that $v \in \mathbf{R}^n$ is a **feasible direction** if $Av = 0$.
- In this case, every point of the form $x + tv$ is also feasible, i.e., $A(x + tv) = b$.
- We say that v is a descent direction for f at x, if for small $t > 0$, $f(x + tv) < f(x)$.
- The Newton step is always a feasible descent direction (except when x is optimal, in which case $\Delta x_{\mathrm{nt}} = 0$).
- Indeed, the second set of equations that define Δx_{nt} are $A\Delta x_{\mathrm{nt}} = 0$, which shows it is a feasible direction; that it is a descent direction follows from

$$\frac{d}{dt} f(x + t\Delta x_{\mathrm{nt}})\bigg|_{t=0} = \nabla f(x)^T \Delta x_{\mathrm{nt}} = -\lambda(x)^2.$$

Affine invariance (1/2)

- The Newton step and decrement for equality constrained optimization are affine invariant.
- Suppose $T \in \mathbf{R}^{n\times n}$ is nonsingular, and define $\bar{f}(y) = f(Ty)$. We have

$$\nabla \bar{f}(y) = T^T \nabla f(Ty), \nabla^2 \bar{f}(y) = T^T \nabla^2 f(Ty) T,$$

  and the equality constraint $Ax = b$ becomes $ATy = b$.
- Now consider the problem of minimizing $\bar{f}(y)$, subject to $ATy = b$.
- The Newton step Δy_{nt} at y is given by the solution of

$$\begin{bmatrix} T^T \nabla^2 f(Ty) T & T^T A^T \\ AT & 0 \end{bmatrix} \begin{bmatrix} \Delta y_{\mathrm{nt}} \\ \bar{w} \end{bmatrix} = \begin{bmatrix} -T^T \nabla f(Ty) \\ 0 \end{bmatrix}.$$

Affine invariance (2/2)

- Comparing with the Newton step Δx_{nt} for f at $x = Ty$, given in

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

we see that $T\Delta y_{\mathrm{nt}} = \Delta x_{\mathrm{nt}}$ (and $w = \bar{w}$), i.e., the Newton steps at y and x are related by the same change of coordinates as $Ty = x$.

Newton's method for equality constrained minimization

- **Algorithm 10.1** Newton's method for equality constrained minimization.

  given starting point $x \in \textbf{dom}\ f$ with $Ax = b$, tolerance $\epsilon > 0$.

  repeat

  1. Compute the Newton step and decrement $\Delta x_{\rm nt}, \lambda(x)$.
  2. Stopping criterion. **quit** if $\lambda^2/2 \le \epsilon$.
  3. Line search. Choose step size t by backtracking line search.
  4. Update. $x := x + t\Delta x_{\rm nt}$.

- The method is called a **feasible descent method**, since all the iterates are feasible, with $f(x^{(k+1)}) < f(x^{(k)})$ (unless $x^{(k)}$ is optimal).
- Newton's method requires that the KKT matrix be invertible at each x.

Newton's method and elimination (1/5)

- It can be shown that the iterates in Newton's method for the equality constrained problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b, \end{array}$$

coincide with the iterates in Newton's method applied to the reduced problem

$$\text{minimize} \quad \tilde{f}(z) = f(Fz + \hat{x}),$$

- Suppose F satisfies $\mathcal{R}(F) = \mathcal{N}(A)$ and rank $F = n - p$, and $\hat{x}$ satisfies $A\hat{x} = b$.

Newton's method and elimination (2/5)

- The gradient and Hessian of the reduced objective function $\tilde{f}(z) = f(Fz + \hat{x})$ are

$$\nabla \tilde{f}(z) = F^T \nabla f(Fz + \hat{x}), \quad \nabla^2 \tilde{f}(z) = F^T \nabla^2 f(Fz + \hat{x}) F.$$

- From the Hessian expression, we see that the Newton step for the equality constrained problem is defined, i.e., the KKT matrix

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}$$

is invertible, if and only if the Newton step for the reduced problem is defined, i.e., $\nabla^2 \tilde{f}(z)$ is invertible.

Newton's method and elimination (3/5)

- The Newton step for the reduced problem is

$$\Delta z_{\mathrm{nt}} = -\nabla^2 \tilde{f}(z)^{-1} \nabla \tilde{f}(z) = -(F^T \nabla^2 f(x) F)^{-1} F^T \nabla f(x),$$

  where $x = Fz + \hat{x}$.

- This search direction for the reduced problem corresponds to the direction

$$F \Delta z_{\mathrm{nt}} = -F(F^T \nabla^2 f(x) F)^{-1} F^T \nabla f(x)$$

  for the original, equality constrained problem.

- We claim that $\Delta x_{\mathrm{nt}} = F \Delta z_{\mathrm{nt}}$.

Newton's method and elimination (4/5)

- To show this, we take $\Delta x_{\rm nt} = F \Delta z_{\rm nt}$, choose $w = -(AA^T)^{-1} A(\nabla f(x) + \nabla^2 f(x) \Delta x_{\rm nt})$, and verify that the equations defining the Newton step,

$$\nabla^2 f(x) \Delta x_{\rm nt} + A^T w + \nabla f(x) = 0, \quad A \Delta x_{\rm nt} = 0,$$

hold.
- The second equation, $A \Delta x_{\rm nt} = 0$, is satisfied because $AF = 0$. To verify the first equation, we observe that

$$\begin{aligned} & \begin{bmatrix} F^T \\ A \end{bmatrix} \left(\nabla^2 f(x) \Delta x_{\rm nt} + A^T w + \nabla f(x) \right) \\ = & \begin{bmatrix} F^T \nabla^2 f(x) \Delta x_{\rm nt} + F^T A^T w + F^T \nabla f(x) \\ A \nabla^2 f(x) \Delta x_{\rm nt} + AA^T w + A \nabla f(x) \end{bmatrix} \\ = & \ 0. \end{aligned}$$

Newton's method and elimination (5/5)

- Since the matrix on the left of the first line is nonsingular, we conclude that the conditions

$$\nabla^2 f(x)\Delta x_{\mathrm{nt}} + A^T w + \nabla f(x) = 0, \quad A\Delta x_{\mathrm{nt}} = 0,$$

hold.

- In a similar way, the Newton decrement $\tilde{\lambda}(z)$ of $\tilde{f}$ at z and the Newton decrement of f at x turn out to be equal:

$$\begin{aligned}
\tilde{\lambda}(z)^2 &= \Delta z_{\mathrm{nt}}^T \nabla^2 \tilde{f}(z) \Delta z_{\mathrm{nt}} \\
&= \Delta z_{\mathrm{nt}}^T F^T \nabla^2 f(x) F \Delta z_{\mathrm{nt}} \\
&= \Delta x_{\mathrm{nt}}^T \nabla^2 f(x) \Delta x_{\mathrm{nt}} \\
&= \lambda(x)^2.
\end{aligned}$$

Newton step at infeasible points (1/4)

- The Newton's method described in the previous section is a feasible descent method.
- We now describe a generalization of Newton's method that works with infeasible initial points and iterates.
- Starting with the optimality conditions for the equality constrained minimization problem:

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

and assuming $x \in \textbf{dom } f$ as the current point (not necessarily feasible), our goal is to find a step Δx so that $x + \Delta x$ satisfies (at least approximately) the optimality conditions, i.e.,

$$x + \Delta x \approx x^*.$$

Newton step at infeasible points (2/4)

- To do this we substitute $x + \Delta x$ for x^* and w for ν^* in the optimality conditions, and use the first-order approximation

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

for the gradient. Then, we obtain

$$A(x + \Delta x) = b, \nabla f(x) + \nabla^2 f(x) \Delta x + A^T w = 0.$$

- This is a set of linear equations for Δx and w,

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

Newton step at infeasible points (3/4)

- The equations

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

are the same as the equations that define the Newton step at a feasible point x, i.e.,

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

with only one difference: the second block component of the righthand side contains $Ax - b$, which is the residual vector for the linear equality constraints.

Newton step at infeasible points (4/4)

- When x is feasible, the residual vanishes, and the equations reduce to the equations that define the standard Newton step at a feasible point x

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

Interpretation as primal-dual Newton step (1/4)

- The equations

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

  can be interpreted in terms of a primal-dual method for the equality constrained problem.

- This means that we update both the primal variable x, and the dual variable ν, in order to (approximately) satisfy the optimality conditions.
- We express the optimality conditions as $r(x^*, \nu^*) = 0$, where $r : \mathbf{R}^n \times \mathbf{R}^p \to \mathbf{R}^n \times \mathbf{R}^p$ is defined as

$$r(x, \nu) = (r_{\mathrm{dual}}(x, \nu), r_{\mathrm{pri}}(x, \nu)).$$

Interpretation as primal-dual Newton step (2/4)

- Here

$$r_{\mathrm{dual}}(x,\nu) = \nabla f(x) + A^T \nu, \quad r_{\mathrm{pri}}(x,\nu) = Ax - b$$

are the **dual residual** and **primal residual**, respectively.

- The first-order Taylor approximation of r, near our current estimate y, is

$$r(y+z) \approx \hat{r}(y+z) = r(y) + Dr(y)z,$$

where $Dr(y) \in \mathbf{R}^{(n+p)\times(n+p)}$ is the derivative of r at y.

- We define the **primal-dual Newton step** Δy_{pd} as the step z for which the Taylor approximation $\hat{r}(y+z)$ vanishes, i.e.,

$$Dr(y)\Delta y_{\mathrm{pd}} = -r(y).$$

Interpretation as primal-dual Newton step (3/4)

- Note that
$$\Delta y_{\mathrm{pd}} = (\Delta x_{\mathrm{pd}}, \Delta \nu_{\mathrm{pd}})$$
gives both a primal and a dual step.
- Then, we can express $Dr(y)\Delta y_{\mathrm{pd}} = -r(y)$ as
$$\begin{aligned}\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} \Delta x_{\mathrm{pd}} \\ \Delta \nu_{\mathrm{pd}} \end{bmatrix} &= -\begin{bmatrix} r_{\mathrm{dual}} \\ r_{\mathrm{pri}} \end{bmatrix} \\ &= -\begin{bmatrix} \nabla f(x) + A^T \nu \\ Ax - b \end{bmatrix}.\end{aligned}$$
- Writing $\nu + \Delta \nu_{\mathrm{pd}}$ as ν^+, we can express this as
$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} \Delta x_{\mathrm{pd}} \\ \nu^+ \end{bmatrix} = -\begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix}$$
which is exactly the same set of equations as that in page 42.

Interpretation as primal-dual Newton step (4/4)

- Therefore, we have $\Delta x_{\mathrm{nt}} = \Delta x_{\mathrm{pd}}$, $w = \nu^+ = \nu + \Delta\nu_{\mathrm{pd}}$, which shows that the (infeasible) Newton step is the same as the primal part of the primal-dual step, and the associated dual vector w is the updated primal-dual variable $\nu^+ = \nu + \Delta\nu_{\mathrm{pd}}$.
- The expression for the Newton step and dual step

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{pd}} \\ \Delta \nu_{\mathrm{pd}} \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + A^T \nu \\ Ax - b \end{bmatrix},$$

and that for the Newton step and dual variable

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

are equivalent, but each reveals a different feature of the Newton step.

Residual norm reduction property

- The Newton direction, at an infeasible point, is not necessarily a descent direction for f.
- But it can be shown that the norm of the residual decreases in the Newton direction:

$$\frac{d}{dt}||r(y + t\Delta y_{\mathrm{pd}})||_2^2\bigg|_{t=0} = 2r(y)^T Dr(y)\Delta y_{\mathrm{pd}} = -2r(y)^T r(y),$$

which implies

$$\frac{d}{dt}||r(y + t\Delta y_{\mathrm{pd}})||_2\bigg|_{t=0} = -||r(y)||_2.$$

Full step feasibility property (1/3)

- The Newton step Δx_{nt} defined above has the property (by construction) that

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

  So, if a step length of one is taken using the Newton step Δx_{nt}, the following iterate will be feasible.
- Once x is feasible, the Newton step becomes a feasible direction, so all future iterates will be feasible, regardless of the step sizes taken.
- More generally, we can analyze the effect of a damped step on the equality constraint residual r_{pri}.

Full step feasibility property (2/3)

- With a step length $t \in [0, 1]$, the next iterate is $x^+ = x + t\Delta x_{\mathrm{nt}}$, so the equality constraint residual at the next iterate is

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

- Thus, a damped step, with length t, causes the residual to be scaled down by a factor $1 - t$. Now suppose that we have

$$x^{(i+1)} = x^{(i)} + t^{(i)} \Delta x_{\mathrm{nt}}^{(i)},$$

for $i = 0, ..., k - 1$, where $\Delta x_{\mathrm{nt}}^{(i)}$ is the Newton step at the point $x^{(i)} \in \textbf{dom}\ f$, and $t^{(i)} \in [0, 1]$.

Full step feasibility property (3/3)

- Then we have

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

where $r^{(i)} = Ax^{(i)} - b$ is the residual of $x^{(i)}$. This formula shows that the primal residual at each step is in the direction of the initial primal residual, and is scaled down at each step.

- It also shows that once a full step is taken, all future iterates are primal feasible.

Infeasible start Newton method (1/3)

- **Algorithm 10.2** Infeasible start Newton method.
  given starting point $x \in \textbf{dom } f, \nu$, tolerance $\epsilon > 0, \alpha \in (0, 1/2), \beta \in (0, 1)$.
  repeat
  1. Compute primal and dual Newton steps $\Delta x_{\text{nt}}, \Delta \nu_{\text{nt}}$.
  2. *Backtracking line search on* $||r||_2$.
     $t := 1$.
     while $||r(x + t\Delta x_{\text{nt}}, \nu + t\Delta \nu_{\text{nt}})||_2 > (1 - \alpha t)||r(x, \nu)||_2$,
     $t := \beta t$.
  3. *Update*. $x := x + t\Delta x_{\text{nt}}, \nu := \nu + t\Delta \nu_{\text{nt}}$.

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

Infeasible start Newton method (2/3)

- This algorithm is very similar to the standard Newton method with feasible starting point, with a few exceptions.
    1. The search directions include the extra correction terms that depend on the primal residual.
    2. The line search is carried out using the norm of the residual, instead of the function value f.
    3. The algorithm terminates when primal feasibility has been achieved, and the norm of the (dual) residual is small.
- Some comments on the line search in step 2:
    - Using the norm of the residual in the line search can increase the cost, compared to a line search based on the function value, but the increase is usually negligible.
    - The line search must terminate in a finite number of steps, since the line search exit condition is satisfied for small t:

$$\frac{d}{dt}||r(y + t\Delta y_{\mathrm{pd}})||_2\Big|_{t=0} = -||r(y)||_2.$$

Infeasible start Newton method (3/3)

- Comments on conditions for the iterate to become feasible.
  - The equation $A(x + \Delta x_{\mathrm{nt}}) = b$ shows that if at some iteration the step length is chosen to be one, the next iterate will be feasible.
  - Thereafter, all iterates will be feasible, and therefore the search direction for the infeasible start Newton method coincides, once a feasible iterate is obtained, with the search direction for the feasible Newton method.