Computational Method of Optimization

Constrained Optimization- Lecture X

Constrained Optimization

- Agenda

- equality constrained minimization

- eliminating equality constraints

- Newton’s method with equality constraints

- infeasible start Newton method

Source: Convex Optimization — Boyd & Vandenberghe

Equality constrained minimization

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

- f convex, twice continuously differentiable
- $A \in \mathbf{R}^{p \times n}$ with $\mathbf{rank}\, A = p$
- we assume $p^\star$ is finite and attained

optimality conditions: $x^\star$ is optimal iff there exists a $\nu^\star$ such that

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

This is going to turn out to be not much different from the unconstrained case. If we want to minimize f(x) subject to Ax = b. We'll assume A is large, and it's full rank, in particular Ax = b is solvable. And we'll also assume that the problem is bounded below, and f is convex and continuous is twice differentiable, a continuous second derivative. And the optimality conditions are simple. It's if and only if there exists a ν, such that $\nabla f(x^\star)+A^T\nu^\star=0$. So that's the classic Lagrange's multiplying rule. However, in this case it's if and only if, right. There are no constrained qualifications of any kind here. These are the necessary conditions.

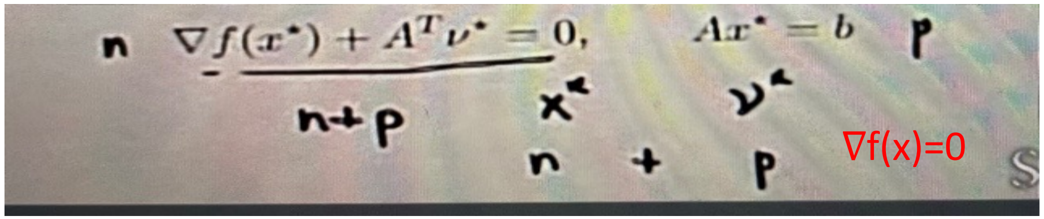
Equality constrained minimization

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

- f convex, twice continuously differentiable
- $A \in \mathbf{R}^{p \times n}$ with $\mathbf{rank}\, A = p$
- we assume $p^\star$ is finite and attained

optimality conditions: $x^\star$ is optimal iff there exists a $\nu^\star$ such that

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



Another way to think of it is this, is we want to solve a set of n + p equations in the variables x^* and $\nu^\star$, which just coincidentally happen to have dimension n + p. Because this is n equations and that is p equations. When you do unconstrained minimization of f, it is to solve the n nonlinear equations and n variables; nonlinear, unless f is quadratic, in which case this is linear. But in general, nonlinear equations; n equations and unknowns. That's what Newton's method is doing for example.

In fact, you want to solve n + p equations and n + p unknowns, but there's some structure here. For example, the p, the added p equations are linear. These are the nonlinear ones. And they are linear in the v.

Equality constrained convex quadratic minimization

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

optimality condition:

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ \nu^\star \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix} \begin{array}{l} \text{dual feas.} \\ \text{primal feas.} \end{array}$$

- coefficient matrix is called KKT matrix
- KKT matrix is nonsingular if and only if

$$Ax = 0, \quad x \neq 0 \qquad \Longrightarrow \qquad x^T P x > 0$$

- equivalent condition for nonsingularity: $P + A^T A \succ 0$

If you minimize a quadratic subject to an equality constraint, this is a positive semi-definite. Then you write the gradient. The gradient is $Px^\star + q + A^T v^\star = 0$. It's a set of linear equations, which you write this way.

This set of n + p linear equations in the n + p variables x⋆, ν⋆ is called the Karush-Kuhn-Tucker (KKT) system for the equality constrained quadratic optimization problem. The coefficient matrix is called the KKT matrix.

It looks like this. Its bottom is a zero. Here, it's got an A and an A^T here, and it has a positive, we'll assume P positive definite. So that's the so called KKT matrix.

Equality constrained convex quadratic minimization

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

optimality condition:

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ \nu^\star \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix} \begin{array}{l} \text{dual feas.} \\ \text{primal feas.} \end{array}$$

- coefficient matrix is called KKT matrix
- KKT matrix is nonsingular if and only if

$$Ax = 0, \quad x \neq 0 \qquad \Longrightarrow \qquad x^T P x > 0$$

- equivalent condition for nonsingularity: $P + A^T A \succ 0$

It's non-singular if and only if the following is true. On the null space of A, P is positive definite. That’s the condition here. That's exactly the condition. And that's the same completely equivalent to P + A^TA is positive definite. It says that here A^TA is positive definite. This thing can have– the null space of this is exactly the null space of A. So, what you need is you need P to step in and be positive on the null space of A. That’s why these are equivalent. So this, the so called KKT for minimizing a convex quadratic with equality constraints; it's a set of linear equations. By linear equations already with structure,
It's got this 2 x 2 block structure.

Eliminating equality constraints

represent solution of $\{x \mid Ax = b\}$ as

{{x| A x = b }= x_p+ N(A)}

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

x_p

- $\hat{x}$ is (any) particular solution
- range of $F \in \mathbf{R}^{n \times (n-p)}$ is nullspace of A ($\mathbf{rank}\, F = n - p$ and $AF = 0$)

reduced or eliminated problem

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

- an unconstrained problem with variable $z \in \mathbf{R}^{n-p}$
- from solution $z^\star$, obtain $x^\star$ and $\nu^\star$ as

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

You can eliminate equality constraints; You write the solution set of a set of linear equalities, this is in so called constraint form. Because it's a set of x that satisfies some equations. You can also write it in free parameter form. it's $\hat{x}$ or sometimes x_p, collides with that one. it's a particular solution of Ax = b, plus an element of the null space.
Obviously, this is the case, set of x's to the Ax equals b, is equal to x particular plus the null space of A. So that's the condition.

Eliminating equality constraints

represent solution of $\{x \mid Ax = b\}$ as {{x| A x = b }= x_p+ N(A)}

$$\{x \mid Ax = b\} = \{F z + \underbrace{\hat{x}}_{x_p} \mid z \in \mathbf{R}^{n-p}\}$$

- $\hat{x}$ is (any) particular solution
- range of $F \in \mathbf{R}^{n \times (n-p)}$ is nullspace of A ($\mathbf{rank}\, F = n - p$ and $AF = 0$)

reduced or eliminated problem

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

- an unconstrained problem with variable $z \in \mathbf{R}^{n-p}$
- from solution $z^\star$, obtain $x^\star$ and $\nu^\star$ as

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

Now here to write this out more explicitly, you calculate a matrix F, whose columns span the null space of A, then this would give you this free parameter representation. Now you change variables, and your equality constraint problem becomes this unconstrained problem. Because now you have a free parameter representation of any x that satisfied Ax = b.

It turns out, if you really wanted to solve this, if you had to write a code, for example that solves this, or, or solves the original problem here, that your code must return v^* as well as x^*.

Example

example: optimal allocation with resource constraint

$$\begin{array}{ll} \text{minimize} & f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n) \\ \text{subject to} & x_1 + x_2 + \cdots + x_n = b \end{array}$$

eliminate $x_n = b - x_1 - \cdots - x_{n-1}$, *i.e.*, choose

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

reduced problem:

$$\text{minimize} \quad f_1(x_1) + \cdots + f_{n-1}(x_{n-1}) + f_n(b - x_1 - \cdots - x_{n-1})$$

(variables $x_1, \ldots, x_{n-1}$)

Let's look at an example here. The optimal allocation with resource restraint looks like this. So you minimize a sum. Basically, I have a set of resources, that's our total amount of resource, that's b, and I divide that up among n users of the resource. And I want to minimize some cost of this thing. But the costs are separate. So I want to minimize $f_1(x)$, ..and $f_n(x)$, each person has a different cost function. And we run an allocation that sort of minimizes the total cost across the population while satisfying resource constraint.

Example

example: optimal allocation with resource constraint

$$\begin{array}{ll} \text{minimize} & f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n) \\ \text{subject to} & x_1 + x_2 + \cdots + x_n = b \end{array}$$

eliminate $x_n = b - x_1 - \cdots - x_{n-1}$, *i.e.*, choose

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

reduced problem:

$$\text{minimize} \quad f_1(x_1) + \cdots + f_{n-1}(x_{n-1}) + f_n(b - x_1 - \cdots - x_{n-1})$$

(variables $x_1, \ldots, x_{n-1}$)

Imagine x_1 though $x_{(n-1)}$ is free variables, once you've fixed these, then x_n is simply $b - x_1 -..x_{n-1}$ like that. By the way, this is the same as choosing F to be this matrix here, and $\hat{x}$ is be_n here. The particular solutions give all the resources to the nth user. It is a feasible point.

The reduced problem looks like this, which we could then minimize now.

Newton step

Newton step Δx_{nt} of f at feasible x is given by solution v of

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

interpretations

- Δx_{nt} solves second order approximation (with variable v)

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

- Δx_{nt} equations follow from linearizing optimality conditions

$$\nabla f(x+v) + A^T w \approx \nabla f(x) + \nabla^2 f(x) v + A^T w = 0, \qquad A(x+v) = b$$

Newton's method to problems with equality constraints. That looks like this. You look at the Newton step Δx_{nt} , and you write down the KKT conditions where you put in the local, the Hessian. That's the KKT matrix. That's a step in your x and w is a dual variable.

You can see that without equality constraints it reduces to the Newton step Δx_{nt} (with v approximation), because it's simple the Hessian times the gradient, Hessian times the Newton is equal to negative gradient. That's Newton.

Newton step

Newton step Δx_{nt} of f at feasible x is given by solution v of

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

interpretations

Min$_x$ f(x)
with Ax=b

- Δx_{nt} solves second order approximation (with variable v)

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

Av=0

- Δx_{nt} equations follow from linearizing optimality conditions

$$\nabla f(x+v) + A^T w \approx \nabla f(x) + \nabla^2 f(x) v + A^T w = 0, \qquad A(x+v) = b$$

Here are some interpretations. One is this. It's completely natural. It says, I'm solving this problem, minimize f(x) subject to Ax = b. you have a current x, which is a guess. And it's feasible, so it satisfies Ax = b. We'll develop f into a second order Taylor series at current point, and we will get this thing. Now that right there is a quadratic function of v.

Instead of minimizing f, we'll minimize quadratic model, but subject to Ax = b. Ax = b, so Av has to be 0, so v has to be in the null space. That makes perfect sense.

If you are at a point, which is a solution of Ax =b, and you are considering directions to go in where you will still satisfy Ax = b, you have to move into null space. Therefore, it will say Av = 0. And the second one will be a quadratic, and it's exactly this.

Newton step

Newton step Δx_{nt} of f at feasible x is given by solution v of

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

interpretations

$\text{Min}_x\ f(x)$
with Ax=b

- Δx_{nt} solves second order approximation (with variable v)

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

Av=0

- Δx_{nt} equations follow from linearizing optimality conditions

$$\nabla f(x+v) + A^T w \approx \nabla f(x) + \nabla^2 f(x) v + A^T w = 0, \qquad A(x+v) = b$$

Another way to do it, to interpret the Newton step is this, is you look at the optimality conditions, which is this here, and which looks like this.

In fact, if we could find v and it satisfied this, because x + v would be optimal.
The problem is that this here is nonlinear.
So what we'll do is we'll have an affine approximation of this, and we're going to call that the ∇f (x) +∇²f (x)v.
Now that's the Newton step for equality constraints.

Newton decrement

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

properties

- gives an estimate of $f(x) - p^\star$ using quadratic approximation $\widehat{f}$:

$$f(x) - \inf_{Ay=b} \widehat{f}(y) = \frac{1}{2}\lambda(x)^2$$

- directional derivative in Newton direction:

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

- in general, $\lambda(x) \neq \left(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)\right)^{1/2}$

Now in this case, there's a Newton decrement
½ $\lambda(x)^2$ is your predicted decrease in f
By the way, this tells you that if λ is small, then your decrease in f is going to be small.
It predicts a small decrease

Newton's method with equality constraints

given starting point $x \in \mathbf{dom}\, f$ with $Ax = b$, tolerance $\epsilon > 0$.
repeat
1. Compute the Newton step and decrement Δx_{nt}, $\lambda(x)$.
2. *Stopping criterion.* **quit** if $\lambda^2/2 \leq \epsilon$.
3. *Line search.* Choose step size t by backtracking line search.
4. *Update.* $x := x + t\Delta x_{\mathrm{nt}}$.

- a feasible descent method: $x^{(k)}$ feasible and $f(x^{(k+1)}) < f(x^{(k)})$
- affine invariant

Newton's method is for equality constraints. You start with a point that satisfies Ax = b, and it's in the domain of f, and you compute to do the step in the Newton decrement. You quit if the Newton decrement is small enough, otherwise you do a line search. Backtracking is fine, and then you update. It's a feasible descent method. It's affined in variant, so you can change coordinates. It makes absolutely no difference.

Newton's method and elimination

Newton's method for reduced problem

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

- variables $z \in \mathbf{R}^{n-p}$
- $\hat{x}$ satisfies $A\hat{x} = b$; $\mathbf{rank}\, F = n - p$ and $AF = 0$
- Newton's method for $\tilde{f}$, started at $z^{(0)}$, generates iterates $z^{(k)}$

Newton's method with equality constraints

when started at $x^{(0)} = Fz^{(0)} + \hat{x}$, iterates are

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

hence, don't need separate convergence analysis

Newton's method this way is identical to Newton's method with elimination. You get the same iterates, therefore everything we know, self-concordance, classical analysis, a practical performance, such as you've found out a little bit about so far, it's all the same.

Newton step at infeasible points

2nd interpretation of slide 12 extends to infeasible x (i.e., Ax≠ b)

linearizing optimality conditions at infeasible x (with $x \in \textbf{dom}\, f$) gives

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{nt}} \\ w \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix} \qquad (1)$$

primal-dual interpretation

- write optimality condition as $r(y) = 0$, where

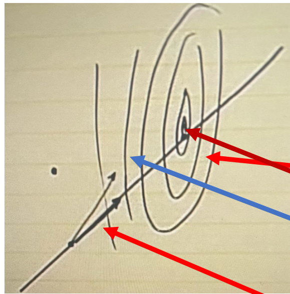
$$y = (x, \nu), \qquad r(y) = (\nabla f(x) + A^T \nu, Ax - b)$$

- linearizing $r(y) = 0$ gives $r(y + \Delta y) \approx r(y) + Dr(y)\Delta y = 0$:

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

  same as (1) with $w = \nu + \Delta \nu_{\mathrm{nt}}$

Newton method for infeasible points. It says, suppose, it's going to lead to an infeasible method. You have the idea of a Newton step even when you are not in the feasible set. It is even just explained by drawing a picture,



Here's your equality constraints, and let's say your function looks like this.

Here are some level sets of function or something like that. So that's the optimal point. So here are some level sets.

Now if you are here, the Newton direction, we should have drawn it more skewed so the center was over there.

Newton step at infeasible points

2nd interpretation of slide 12 extends to infeasible x (i.e., $Ax \neq b$)

linearizing optimality conditions at infeasible x (with $x \in \mathbf{dom}\, f$) gives

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\rm nt} \\ w \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix} \qquad (1)$$

primal-dual interpretation

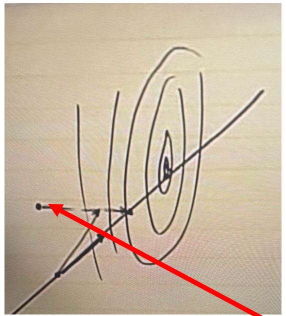
- write optimality condition as $r(y) = 0$, where

$$y = (x, \nu), \qquad r(y) = (\nabla f(x) + A^T \nu, Ax - b)$$

- linearizing $r(y) = 0$ gives $r(y + \Delta y) \approx r(y) + Dr(y)\Delta y = 0$:

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

  same as (1) with $w = \nu + \Delta \nu_{\rm nt}$



So the Newton direction is probably going to want a point like that, unconstrained, you have something in the null space, and you have to be feasible, so it's going to push you in that direction. you can imagine that you are here. Now when you're here, you're not feasible. The constraints are Ax = b, you don't have Ax = b here, because you're off this thing.

Newton step at infeasible points

2nd interpretation of slide 12 extends to infeasible x (i.e., $Ax \neq b$)

linearizing optimality conditions at infeasible x (with $x \in \mathbf{dom}\, f$) gives

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{nt}} \\ w \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix} \qquad (1)$$

primal-dual interpretation

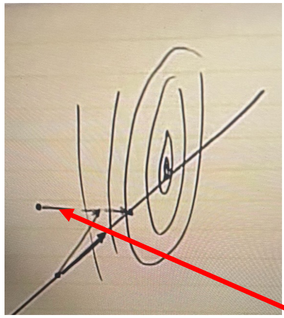
- write optimality condition as $r(y) = 0$, where

$$y = (x, \nu), \qquad r(y) = (\nabla f(x) + A^T \nu, Ax - b)$$

- linearizing $r(y) = 0$ gives $r(y + \Delta y) \approx r(y) + Dr(y)\Delta y = 0$:

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

same as (1) with $w = \nu + \Delta \nu_{\mathrm{nt}}$



Second problem is that you're not optimal here. So you can define a Newton step off here, and it's going to combine two things. One is kind or to approach feasibility, that's No. 1, and No. 2 simultaneously to decrease f. If you are here, you could be in a situation were, feasibility will be there, and you'll have an increase in f.

Once you're feasible, if x were feasible, this would be zero, and this would revert to the standard Newton direction. But if you're not feasible this second part down here, which is a zero when you're feasible, contributes to your direction, and it's a direction that points you towards feasibility.

Newton step at infeasible points

2nd interpretation of slide 12 extends to infeasible x (i.e., Ax ≠b)

linearizing optimality conditions at infeasible x (with $x \in \mathbf{dom}\, f$) gives

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{nt}} \\ w \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix} \qquad (1)$$

primal-dual interpretation

- write optimality condition as $r(y) = 0$, where

$$y = (x, \nu), \qquad r(y) = (\nabla f(x) + A^T \nu, Ax - b)$$

- linearizing $r(y) = 0$ gives $r(y + \Delta y) \approx r(y) + Dr(y)\Delta y = 0$:

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

same as (1) with $w = \nu + \Delta \nu_{\mathrm{nt}}$

The equations are the same as the equations shown in slide 12 that define the Newton step at a feasible point x, with one difference: the second block component of the righthand side contains Ax − b, which is the residual vector for the linear equality constraints. When x is feasible, the residual vanishes, and the equation 1 reduce to the equation of slide 12 that define the standard Newton step at a feasible point x. Thus, if x is feasible, the step Δx defined by equation 1 coincides with the Newton step described here. For this reason, we use the notation Δx_{nt} for the step Δx defined by eqn 1 and refer to it as the Newton step at x, with no confusion.