Learning objectives. By the end of class, you should be able to

1. explain why Newton's method with $Ax = b$ has the same guarantees as without constraints;
2. explain how the infeasible start Newton method reaches $Ax = b$;
3. handle curved constraints with a merit function or an augmented Lagrangian.

1 Recap: Barriers

Last lecture's goal is to solve

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0 \text{ for all } i \in \{1, \dots, m\}. \end{aligned} \tag{Ineq}$$

The barrier method / interior-point method uses Newton's method to solve

$$\min_x \; F_\mu(x) = \frac{f_0(x)}{\mu} - \sum_{i=1}^{m} \log\left(-f_i(x)\right).$$

Progressively decreasing μ, each time using the previous (near-) solution as the new initial point. Since

$$0 \leq f_0(x^\star(\mu)) - f_0^\star \leq m\mu,$$

eventually obtain solution to constrained problem as $\mu \to 0$.

Question 1. Why is it called an interior-point method?

Today's goal is to solve

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f_0(x) \\ \text{s.t.} \quad & h_j(x) = 0 \text{ for all } j \in \{1,\dots,m\}. \end{aligned} \tag{Eq}$$

Of course, this is just an instance of (Ineq)

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f_0(x) \\ \text{s.t.} \quad & \pm\, h_j(x) \le 0 \text{ for all } j \in \{1,\dots,m\}. \end{aligned}$$

However, can't solve this using barrier method / interior-point method.

Question 2. Why can't barrier method solve (Eq) by posing it as (Ineq)?

Question 3. If we can solve (Eq), how do we use it to solve (Ineq) without knowing an interior point?

2 Equality constrained minimization

Let f be strongly convex and twice differentiable. Consider

$$\min_{x \in \mathbb{R}^n} \; f(x) \text{ s.t. } Ax = b.$$

Even though this is *constrained*, we will show that Newton's method solves it as if it were unconstrained.

To see how this is possible, consider the *quadratic* version.

$$\min_{x \in \mathbb{R}^n} \; f_0 + g^T x + \frac{1}{2} x^T H x \text{ s.t. } Ax = b.$$

Question 4. What is the optimality condition for this problem?

2.1 Feasible start

We begin by assuming that we are given a starting point x that satisfies $Ax = b$. This is called a *feasible start*.

Question 5. Suppose x is feasible, $Ax = b$. Then what condition must the step v satisfy for $x_+ = x + v$ to also be feasible?

For a general f, *truncate* its Taylor series at the current point x after the second-order term,

$$f(x+v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v + O(\|v\|^3).$$

Question 6. What is the optimality condition for the quadratic approximation of f subject to $Av = 0$?

Our critical insight for analyzing equality-constrained Newton's method is that

$$v^T \nabla^2 f(x) v = -\nabla f(x)^T v,$$

which is exactly like the unconstrained case. Therefore, the model decreases by $\frac{1}{2} v^T \nabla^2 f(x) v = -\frac{1}{2} \nabla f(x)^T v$.

Question 7. Where does the identity $v^T \nabla^2 f(x) v = -\nabla f(x)^T v$ come from?

Theorem 1. Quadratic convergence with equality constraints. Let f have L_2-Lipschitz Hessian and be μ-strongly convex, and let $(x_\star, z_\star)$ satisfy the KKT conditions. Then from a feasible x, the full-step Newton iteration $x_+ = x + v$, with v from the Newton system, satisfies

$$\|x_+ - x_\star\| \leq \frac{L_2}{2\mu}\|x - x_\star\|^2.$$

Unfortunately, as in Lecture 8, full-step Newton can fail outside its convergence radius.

From a feasible x, compute the Newton search direction v from

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

Given parameters $c_1 \in (0, \frac{1}{2})$ and $c_2 \in (0, 1)$, pick stepsize α by testing $\alpha \in \{1, c_2, c_2^2, c_2^3, \ldots\}$ until the Armijo condition is satisfied

$$f(x + \alpha v) \leq f(x) + c_1 \alpha \nabla f(x)^T v,$$

and then accept $x_+ = x + \alpha v$. Classical analysis is based on the Newton decrement

$$\delta(x) := \sqrt{v^T \nabla^2 f(x) v} = \sqrt{-\nabla f(x)^T v}.$$

Lemma 2. Let f be L_1-Lipschitz gradient, L_2-Lipschitz Hessian, and μ-strongly convex. Define

$$\eta = \frac{\mu^{3/2}}{L_2} \min\{1, 3(1 - 2c_1)\}, \quad \gamma = c_1 \eta^2 \min\left\{1, \frac{2c_2(1 - c_1)\mu}{L_1}\right\}.$$

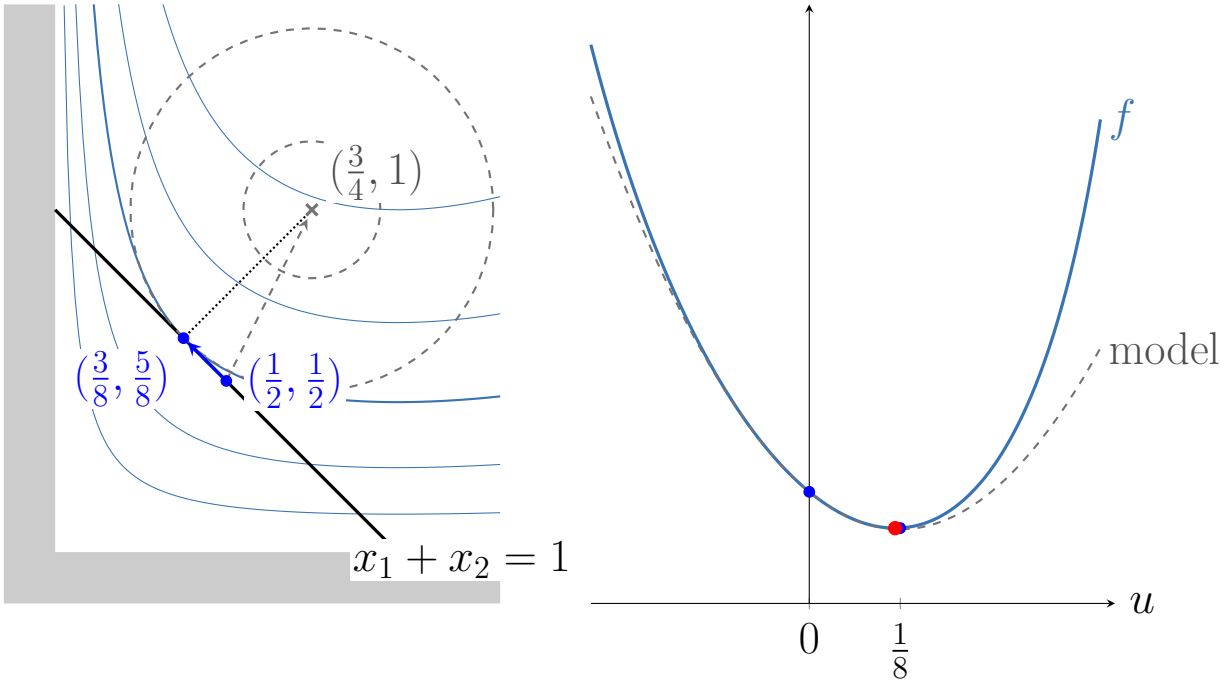
From a feasible x, backtracking Newton has two phases:
(Fixed decrement) If $\delta(x) > \eta$, then $f(x_+) \leq f(x) - \gamma$.
(Quadratic convergence) Once $\delta(x) \leq \eta$, then $\delta(x_+) \leq \frac{1}{2\eta}\delta(x)^2$.

Theorem 3. Under the same conditions as above, backtracking Newton with equality constraints converges to an iterate satisfying $\delta(x_k) \leq \epsilon$ in at most

$$k \leq \left\lceil \frac{f(x_0) - f(x_\star)}{\gamma} + \log_2 \log_2 \left(\frac{2\eta}{\epsilon}\right) \right\rceil \text{ iterations.}$$

Pro: Relatively simple. Completely resolves μ-strongly convex case, with exactly the same constants as Lecture 9.

Con: Needs a feasible start.



Question 8. In the figure, $A = \begin{bmatrix} 1 & 1 \end{bmatrix}$ and $b = 1$. For this specific choice, find Z such that

$$Ax = b \iff x = x_0 + Zu \text{ for some } u \in \mathbb{R}.$$

So the 2D constrained problem is equivalent to the 1D unconstrained problem shown on the right

$$\min_u \ \phi(u) = f(x_0 + Zu).$$

In general, there is always a Z with $AZ = 0$ such that x is feasible if and only if $x = x_0 + Zu$. Our Newton method is identical to Newton's method on $\phi(u)$, even though it never forms Z or u.

2.2 Infeasible start Newton method

For linear constraints $Ax = b$, a feasible start is not hard to find. Nevertheless, it is helpful to have a method that starts from any x in the domain of f.

Question 9. Suppose x is infeasible, $Ax \neq b$. Then what condition must the step v satisfy for $x_+ = x + v$ to be feasible?

Question 10. What is the optimality condition for the quadratic approximation of f subject to $Av = b - Ax$?

However, the critical identity fails. Multiplying the first row by v^T gives

$$v^T H v + (Av)^T z_+ = -g^T v,$$

and now $Av = b - Ax \neq 0$. So $g^T v$ can be positive, and this is not a descent method for f.

Question 11. Let $x_+ = x + \alpha v$ for some $\alpha > 0$. How does $Ax - b$ compare with $A(x_+) - b$?

Lemma 4. Newton direction decreases the residual. Recall the residual

$$r(x, z) = \begin{bmatrix} \nabla f(x) + A^T z \\ Ax - b \end{bmatrix}.$$

If (v, w) solves the Newton system,

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ z_+ \end{bmatrix} = \begin{bmatrix} -g \\ b - Ax \end{bmatrix},$$

then

$$\frac{d}{d\alpha} \left\| r(x + \alpha v,\, z + \alpha w) \right\| \Big|_{\alpha=0} = - \left\| r(x, z) \right\|.$$

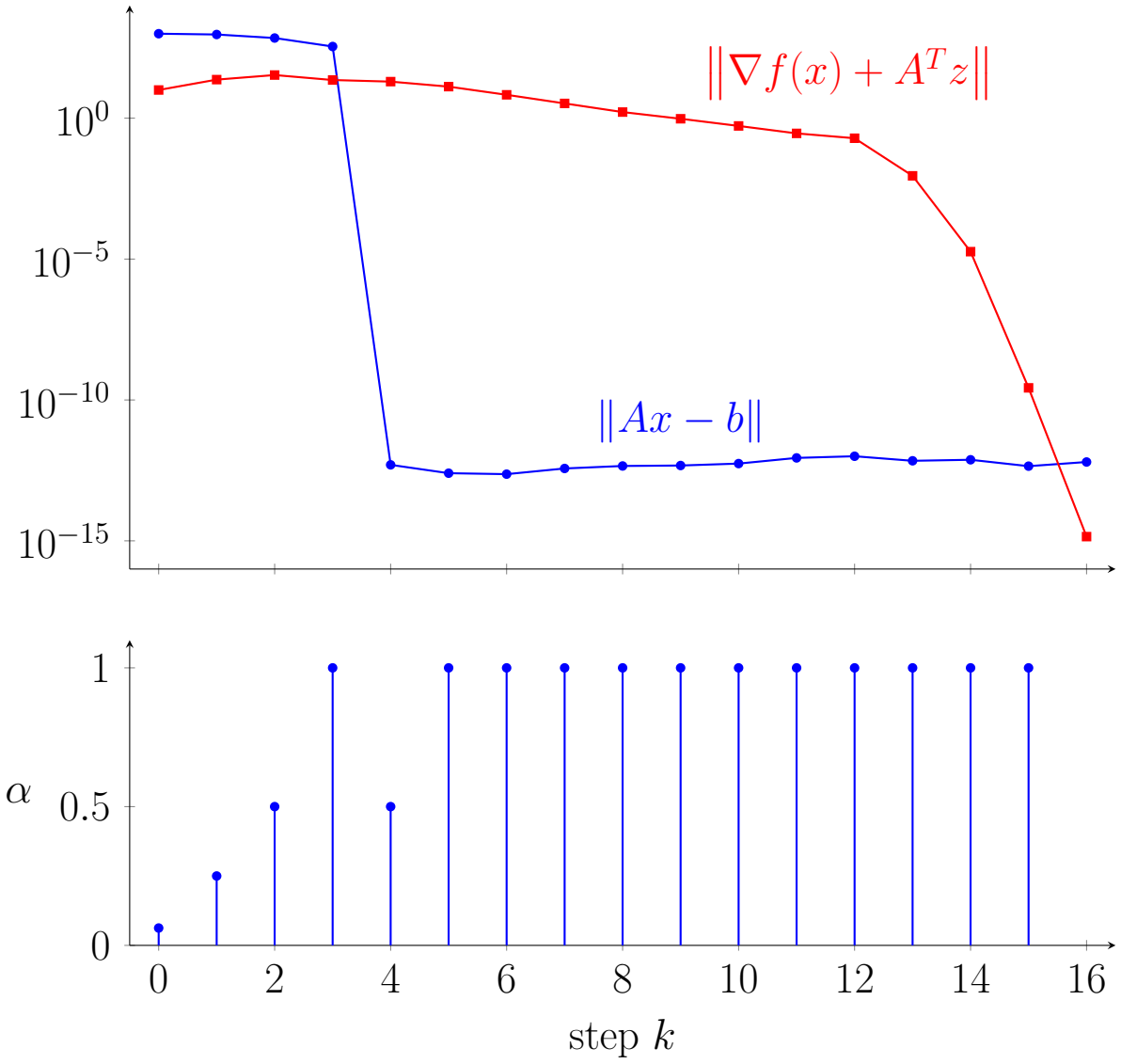
Infeasible start Newton method: Given x in the domain of f, any z, $\epsilon > 0$, $c_1 \in (0, \frac{1}{2})$, and $c_2 \in (0, 1)$, repeat

- **Newton step.** Solve the Newton system for (v, w).
- **Stop** if $Ax = b$ and $\|r(x, z)\| \leq \epsilon$.
- **Backtracking.** Pick stepsize α by testing $\alpha \in \{1, c_2, c_2^2, \ldots\}$ until $x + \alpha v$ is in the domain of f and
$$\|r(x + \alpha v,\, z + \alpha w)\| \leq (1 - c_1\alpha)\, \|r(x, z)\|\,.$$
- **Update** $x \leftarrow x + \alpha v$ and $z \leftarrow z + \alpha w$.

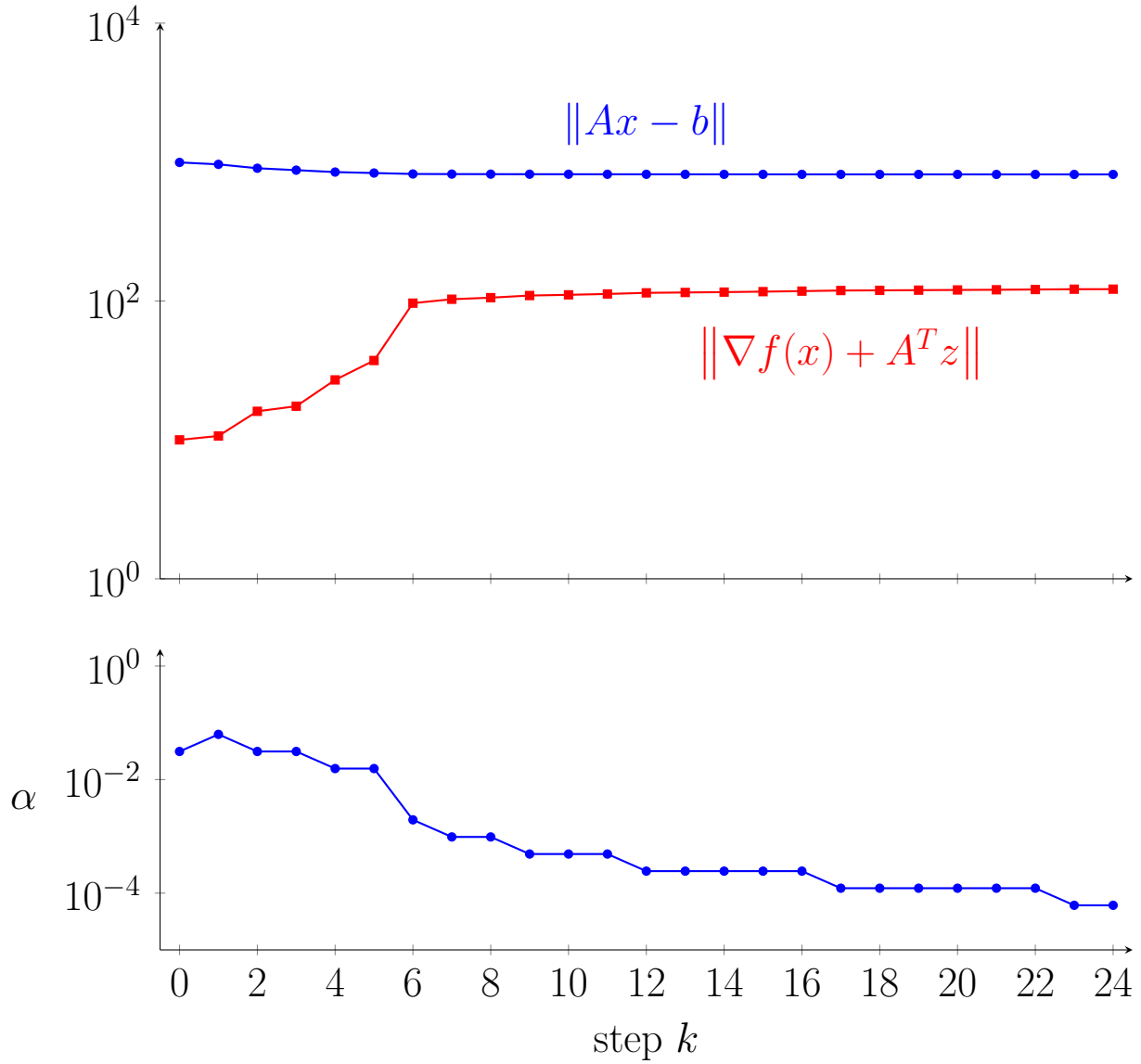
For analytic centering, any $x > 0$ is a valid start, such as $x = 1$ with $z = 0$. No Phase I is needed.

Theorem 5. Infeasible start Newton. Under mild conditions, there are $\eta > 0$ and $\gamma > 0$ such that the infeasible start Newton method has two phases:
(Fixed decrement) If $\|r(x, z)\| > \eta$, then $\|r(x_+, z_+)\| \leq \|r(x, z)\| - \gamma$.
(Quadratic convergence) Once $\|r(x, z)\| \leq \eta$, then $\alpha = 1$ and $\|r(x_+, z_+)\| \leq \frac{1}{2\eta} \|r(x, z)\|^2$.

The first full step lands on $Ax = b$, and from then on this is Newton's method with equality constraints. For the proof, see Boyd and Vandenberghe, Section 10.3.



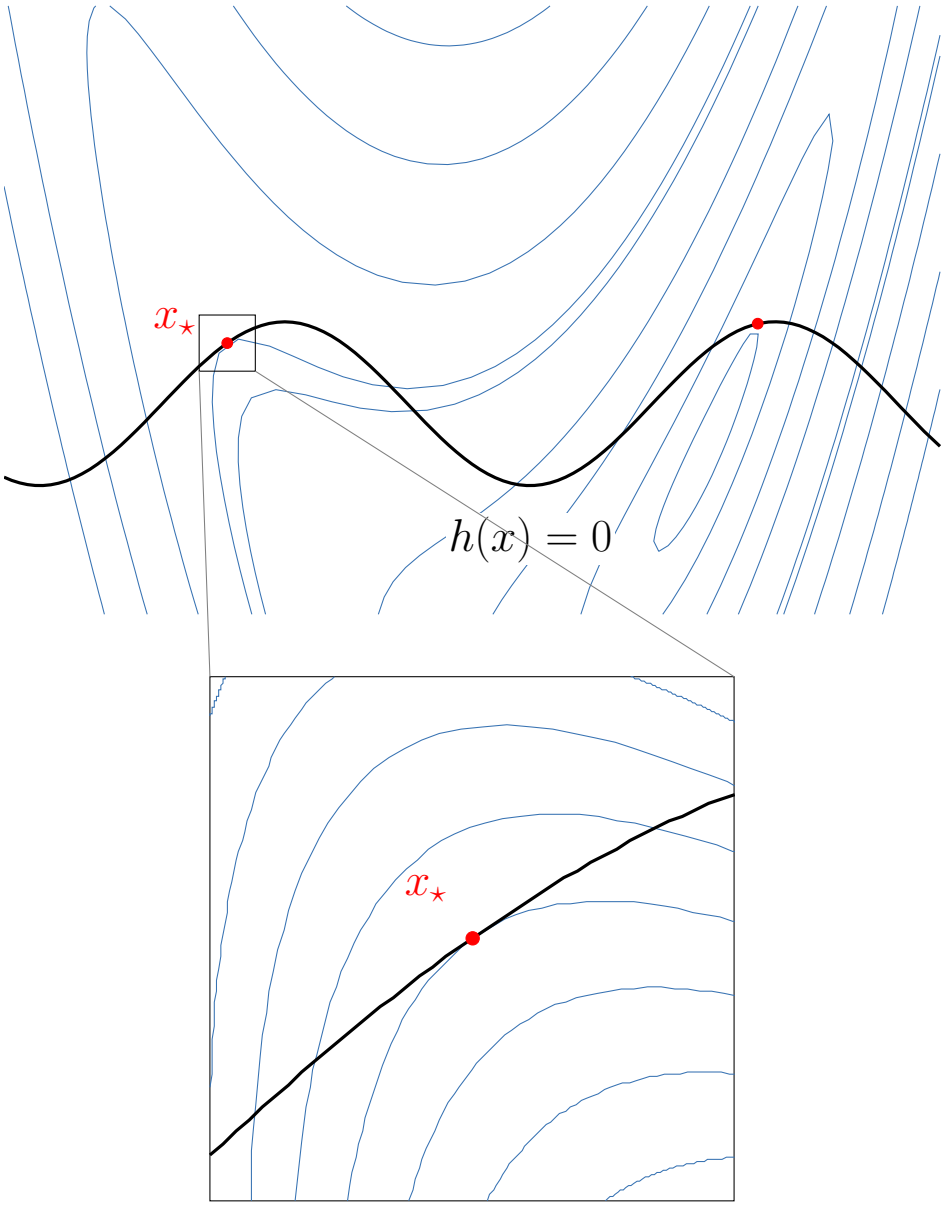
Analytic centering with $n = 100$ and $p = 50$, from $x = 1$ and $z = 0$. Above are the primal residual $\|Ax - b\|$ (blue) and the dual residual $\|\nabla f(x) + A^T z\|$ (red), and below are the stepsizes. The first three steps are damped, with $\alpha = \frac{1}{16}, \frac{1}{4}, \frac{1}{2}$, and $\|Ax - b\|$ shrinks by the factors $\frac{15}{16}, \frac{3}{4}, \frac{1}{2}$. The fourth step is full, so $Ax = b$ from then on, up to rounding. The dual residual decreases slowly, then converges quadratically, 0.19, 9.0×10^{-3}, 1.8×10^{-5}, 2.7×10^{-10}, 1.4×10^{-15}.



The same A, but with its first row made positive and $b_1 = -1$. Then $a_1^T x > 0 > b_1$ for every $x > 0$, so no point in the domain satisfies $Ax = b$. The residuals stall, and the stepsizes shrink toward 0. Unfortunately, the method never reports that the problem is infeasible. It just fails to converge.

3 General equality constraints

The same ideas extend to general equality constraints $h_j(x) = 0$, but the guarantees are weaker.



Above are the level curves of the nonconvex $f_0(x) = (1-x_1)^2+2(x_2-x_1^2)^2$, and the curved constraint $h(x) = 0$. Its two local minimizers are red. Below, zoomed in on $x_\star$, the curve $h = 0$ is nearly straight,

and the level curves of f_0 look like those of a quadratic. This is the picture of a quadratic objective with a linear constraint.

We can no longer impose the constraints exactly on the step. Instead, we linearize each h_j at x,

$$h_j(x+v) \approx h_j(x) + \nabla h_j(x)^T v.$$

Let J be the matrix whose rows are $\nabla h_j(x)^T$, so that $h(x+v) \approx h(x) + Jv$.

Question 12. What condition must the step v satisfy for the linearized constraints to hold at $x+v$?

The second change is that we use the Hessian of the Lagrangian

$$\mathscr{L}(x,z) = f_0(x) + \sum_j z_j h_j(x)$$

in place of $\nabla^2 f$. To see why, write the KKT conditions as $r(x,z) = 0$, with the **residual**

$$r(x,z) = \begin{bmatrix} \nabla f_0(x) + J^T z \\ h(x) \end{bmatrix}.$$

Question 13. Linearize $r(x+v, z+w) \approx 0$ at (x,z). What linear system gives the step (v,w)?

Even a full Newton step can leave the nonlinear constraints violated, so we use a merit function to weigh objective value and constraint violation in the line search.

Merit function: Penalize the violation with a weight $\nu > 0$,

$$\phi_\nu(x) = f_0(x) + \nu \sum_j |h_j(x)|,$$

and apply the Armijo condition to ϕ_ν in place of f_0. If $\nu > |(z_\star)_j|$ for every j, then a strict local minimizer of the problem is also a local minimizer of ϕ_ν.

Augmented Lagrangian: We just used a penalty to choose a step. Here we add a quadratic penalty to the Lagrangian,

$$\mathscr{L}_\nu(x, z) = f_0(x) + \sum_j z_j h_j(x) + \frac{\nu}{2} \sum_j h_j(x)^2.$$

Start with a guess for z. Hold z fixed and minimize over x. Then update $z \leftarrow z + \nu h(x)$ and repeat.

Why this update? At the minimizer we have

$$0 = \nabla_x \mathscr{L}_\nu(x, z) = \nabla f_0(x) + J^T\big(z + \nu h(x)\big).$$

The expression in parentheses is the new z, so the first KKT equation holds at once. The second equation is $h(x) = 0$. As the updates to z go to zero, so does $h(x)$.

Under the usual local convergence conditions, we can keep ν fixed if it is large enough.

4 Summary

The barrier method needs an interior point, and $h_j(x) = 0$ has none. Slack variables turn inequalities into equalities plus $s > 0$, with the barrier on s.

With $Ax = b$, the KKT equations are $\nabla f(x_\star) + A^T z_\star = 0$ and $Ax_\star = b$. For a quadratic f, they are one linear system.

Newton's method with equality constraints minimizes the quadratic model over the affine set, so each step solves one KKT system. The identity $v^T \nabla^2 f(x) v = -\nabla f(x)^T v$ still holds, and its iterates are those of backtracking Newton on $f(x_0 + Zu)$, so the analysis carries over with the same constants. It needs a feasible start.

The infeasible start Newton method starts anywhere in the domain of f. The violation shrinks by the factor $1 - \alpha$, the first full step lands on $Ax = b$ for good, and the Newton direction decreases $\|r\|$, so backtracking on $\|r\|$ replaces the Armijo condition on f. It cannot detect infeasibility.

For curved constraints, linearize h and use the Hessian of the Lagrangian. A full step leaves the constraints, and a merit function or an augmented Lagrangian decides which steps to accept.