Nonlinear Optimization: Algorithms 2: Equality Constrained Optimization

INSEAD, Spring 2006

Jean-Philippe Vert

Ecole des Mines de Paris

Jean-Philippe.Vert@mines.org

Outline

- Equality constrained minimization
- Newton's method with equality constraints
- Infeasible start Newton method

Equality constrained minimization problems

Equality constrained minimization

We consider the problem:

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

- f is supposed to be *convex* and *twice continuously differentiable*.
- A is a $p \times n$ matrix of rank $p < n$ (i.e., fewer equality constraints than variables, and independent equality constraints).
- We assume f^* is finite and attained at x^*

Optimality conditions

Remember that a point $x^* \in \mathbb{R}^n$ is optimal *if and only if* there exists a dual variable $\lambda^* \in \mathbb{R}^p$ such that:

$$\begin{cases} Ax^* & = b \, , \\ \nabla f(x^*) + A^\top \lambda^* & = 0 \, . \end{cases}$$

This is a set of $n + p$ equations in the $n + p$ variables x, λ, called the *KKT system*.

How to solve such problems?

- *Analytically* solve the KKT system (usually not possible)
- *Eliminate the equality constraints* to reduce the constrained problem to an unconstrained problem with fewer variables, and then solve using unconstrained minimization algorithms.
- Solve the *dual problem* using an unconstrained minimization algorithm
- Adapt Newton's methods to the constrained minimization setting (keep the Newton step in the set of feasible directions etc...): *often preferable to other methods*.

Quadratic minimization

Consider the *equality constrained convex quadratic* minimization problem:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}x^\top P x + q^\top x + r \\ \text{subject to} \quad & Ax = b \,, \end{aligned}$$

where $P \in^{n \times n}, P \succeq 0$ and $A \in \mathbb{R}^{p \times n}$. The optimality conditions are:

$$\begin{cases} Ax^* & = b \,, \\ \nabla f(x^*) + A^\top \lambda^* & = 0 \,. \end{cases}$$

Quadratic minimization (cont.)

The optimality conditions can be rewritten as the *KKT system*:

$$\begin{pmatrix} P & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ \lambda^* \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix} .$$

The coefficient matrix in this system is called the *KKT matrix*.

- If the KKT matrix is *nonsingular* (e.g, if $P \succ 0$) there is a *unique optimal* primal-dual pair (x^*, λ^*).
- If the KKT matrix is *singular* but the KKT system *solvable*, any solution yields an optimal pair (x^*, λ^*).
- It the KKT system is *not solvable*, the minimization problem is unbounded below.

Eliminating equality constraints

One general approach to solving the equality constrained minimization problem is to *eliminate* the constraints, and solve the resulting problem with algorithms for unconstrained minimization. The elimination is obtained by a reparametrization of the affine subset:

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

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

Example

Optimal allocation with resource constraint: we want to allocate a single resource, with a fixed total amount b (the budget), to n otherwise independent activities:

$$\begin{aligned} \text{minimize} \quad & f_1(x_1) + f_2(x_2) + \ldots + f_n(x_n) \\ \text{subject to} \quad & x_1 + x_2 + \ldots + x_n = b \, . \end{aligned}$$

Eliminate $x_n = b - x_1 - \ldots - x_{n-1}$, i.e., choose:

$$\hat{x} = be_n \,, \quad F = \begin{pmatrix} I \\ -\mathbf{1}^\top \end{pmatrix} \in \mathbb{R}^{n \times (n-1)} \,,$$

leads to the reduced problem:

$$\min_{x_1, \ldots, x_{n-1}} f_1(x_1) + \ldots + f_{n-1}(x_{n-1}) + f_n(b - x_1 - \ldots - x_{n-1}) \, .$$

Solving the dual

Another approach to solving the equality constrained minimization problem is to solve the *dual*:

$$\max_{\lambda \in \mathbb{R}^p} \left\{ -b^\top \lambda + \inf_x \left\{ f(x) + \lambda^\top A x \right\} \right\} .$$

By hypothesis there is an optimal point so Slater's conditions hold: *strong duality* holds and the dual optimum is attained. If the dual function is twice differentiable, then the methods for unconstrained optimization can be used to maximize it.

Example

The *equality constrained analytic center* is given (for $A \in \mathbb{R}^{p \times n}$) by:

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

The Lagrangian is

$$L(x, \lambda) = -\sum_{i=1}^{n} \log x_i + \lambda^\top (Ax - b)$$

Example (cont.)

We minimize this convex function of x by setting the derivative to 0:

$$\left(A^\top \lambda\right)_i = \frac{1}{x_i}\,,$$

therefore the dual function for $\lambda \in \mathbb{R}^p$ is:

$$q(\lambda) = -b^\top \lambda + n + \sum_{i=1}^{n} \log\left(A^\top \lambda\right)_i$$

We can solve this problem using Newton's method for unconstrained problem, and recover a solution for the primal problem via the simple equation:

$$x_i^* = \frac{1}{\left(A^\top \lambda^*\right)_i}\,.$$

Newton's method with equality constraints

Motivation

Here we describe an *extension of Newton's method* to include linear equality constraint. The methods are almost the same except for two differences:

- the initial point must be *feasible* ($Ax = b$),
- the Newton step must be a *feasible direction* ($A\Delta x_{nt} = 0$).

The Newton step

The *Newton step* of f at a feasible point x for the linear equality constrained problem is given by (the first block of) the solution of:

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

Interpretations

- Δx_{nt} solves the second-order approximation of f at x (with variable v):

$$\begin{aligned} \text{minimize} \quad & f(x) + \nabla f(x)^\top v + \frac{1}{2} v^\top \nabla^2 f(x) v \\ \text{subject to} \quad & A(x+v) = b \,. \end{aligned}$$

The Newton step (cont.)

- When f is exactly quadratic, the Newton update $x + \Delta x_{nt}$ exactly solves the problem and w is the optimal dual variable. When f is nearly quadratic, $x + \Delta x_{nt}$ is a very good approximation of x^*, and w is a good estimate of λ^*.
- *Solution of linearized optimality condition*. Δx_{nt} and w are solutions of the linearized approximation of the optimality condition:

$$\begin{cases} \nabla f\left(x + \Delta x_{nt}\right) + A^\top w & = 0\,, \\ A(x + \Delta x_{nt}) = b\,. \end{cases}$$

Newton decrement

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

- Give an estimate of $f(x) - f^*$ using quadratic approximation:

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

- directional derivative in Newton direction:

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

Newton's method

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

Newton's method and elimination

- *Newton's method for the reduced problem*:

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

starting at $z^{(0)}$, generates iterates $z^{(k)}$.

- *Newton's method with equality constraints*: when started at $x^{(0)} = Fz^{(0)} + \hat{x}$, iterates are:

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

$\Longrightarrow$ the iterates in Newton's method for the equality constrained problem *coincide* with the iterates in Newton's method applied to the unconstrained reduced problem. *All convergence analysis therefore remains valid.*

Summary

- The Newton method for equality constrained optimization problems is the most natural extension of the Newton's method for unconstrained problem: it solves the problem on the affine subset of constraints.
- All results valid for the Newton's method on unconstrained problems remain valid, in particular *it is a good method*.
- Drawback: we need a feasible initial point.

Infeasible start Newton method

Motivation

- Newton's method for constrained problem is a *descent method* that generates a sequence of *feasible points*.
- This requires in particular a feasible point as a starting point.
- Here we generalize Newton's method to work with initial points and iterates that are *not feasible*.
- A price to pay is that it is not necessarily a descent method.

Newton step at infeasible points

The *Newton step* of f at an infeasible point x for the linear equality constrained problem is given by (the first block of) the solution of:

$$\begin{pmatrix} \nabla^2 f(x) & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x_{nt} \\ w \end{pmatrix} = - \begin{pmatrix} \nabla f(x) \\ Ax - b \end{pmatrix} .$$

- When x is feasible, $Ax - b = 0$ and we recover the classical Newton step for equality constrained problems.

Interpretation 1

Remember the optimality conditions:

$$Ax^* = b \; , \quad \nabla f\left(x^*\right) + A^\top \lambda^* = 0 \; .$$

Let x be the current point (not necessarily feasible). Our goal is to find a step Δx s.t. $x + \Delta x$ satisfies approximately the optimality condition. After linearization we get:

$$A\left(x + \Delta x\right) = b \; , \quad \nabla f(x) + \nabla^2 f(x) \Delta x + A^\top w = 0 \; ,$$

i.e., the definition of the Newton step.

Primal-dual interpretation

A *primal-dual method* is a method in which we update both the primal variable x and the dual variable λ, in order to (approximately) satisfy the optimality conditions.
For a given primal-dual pair $y = (x, \lambda)$, the optimality conditions are $r(y) = 0$ with

$$r(y) = \left(\nabla f(x) + A^\top \lambda, Ax - b \right) .$$

Linearizing $r(y) = 0$ gives $r(y + \Delta y) = r(y) + Dr(y)\Delta y = 0$, i.e.:

$$\begin{pmatrix} \nabla^2 f(x) & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x_{nt} \\ \Delta \lambda_{nt} \end{pmatrix} = - \begin{pmatrix} \nabla f(x) + A^\top \lambda \\ Ax - b \end{pmatrix} .$$

which is similar to the Newton step with $w = \lambda + \Delta \lambda_{nt}$.

Residual norm

The Newton direction is not necessarily a descent direction:

$$
\begin{aligned}
\frac{d}{dt} f\left(x + t\Delta x\right)|_{t=0} &= \nabla f(x)^\top \Delta x \\
&= -\Delta x^\top \left(\nabla^2 f(x)\Delta x + A^\top w\right) \\
&= -\Delta x^\top \nabla^2 f(x)\Delta x + (Ax - b)^\top w \, ,
\end{aligned}
$$

which is not necessarily negative (unless $Ax = b$). The *residual* of the primal-dual interpretation, however decreases in norm at each iteration because:

$$
\frac{d}{dt}\| r\left(y + t\Delta y_{pd}\right) \||_{t=0} = -\| r(y) \|_2 \leq 0 \, ,
$$

therefore the norm $\| r \|_2$ can be used to *measure the progress* of the infeasible start Newton method, for example in the line search.

Infeasible start Newton method

- *given* starting point $x \in \mathbb{R}^n$, tolerance $\epsilon > 0, \alpha \in (0, 1/2), \beta \in (0, 1)$
- *repeat*
  1. Compute primal and dual Newton steps $\Delta x_{nt}, \Delta \lambda_{nt}$
  2. Backtracking line search on $\| r \|_2$:
     - t:=1
     - *while* $\| r \left(x + t\Delta x_{nt}, \lambda + t\Delta \lambda_{nt} \right) \|_2 > (1 - \alpha t)\| r(x, \lambda) \|_2 , \quad t = \beta t$.
  3. Update: $x = x + t\Delta x_{nt}, \lambda = \lambda + t\Delta \lambda_{nt}$.

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

Example

Equality constrained analytic centering:

$$\begin{aligned} \text{minimize} \quad & -\sum_{i=1}^{n} \log x_i \\ \text{subject to} \quad & Ax = b \,. \end{aligned}$$

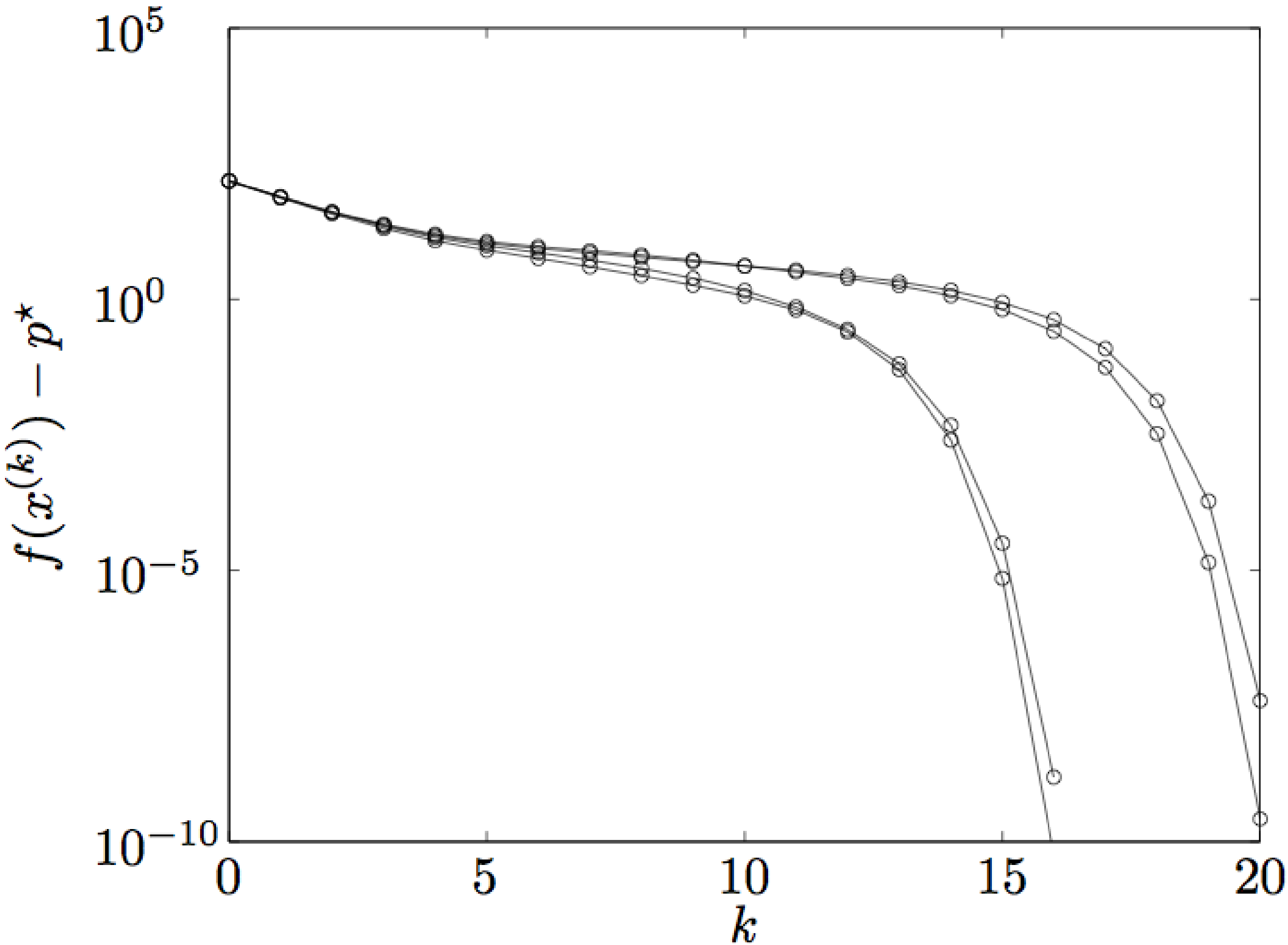
The dual problem is

$$\max_{\lambda} -b^\top \lambda + \sum_{i=1}^{n} \log \left(A^\top \lambda\right)_i + n \,.$$

We compare *three methods* for solving this problem with $A \in \mathbb{R}^{100 \times 500}$, with different starting points.

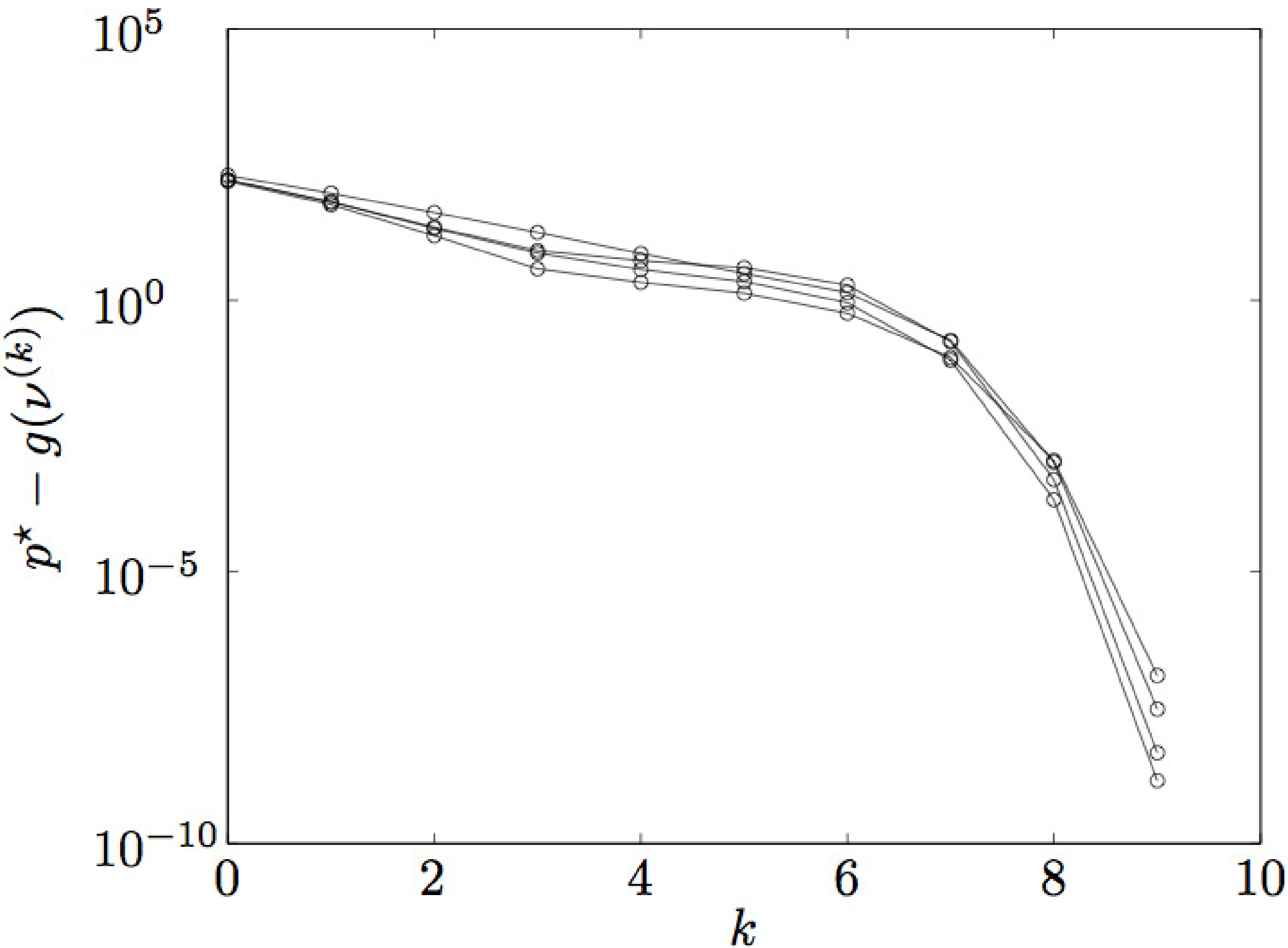
Example (cont)

1. Newton's method with equality constraint



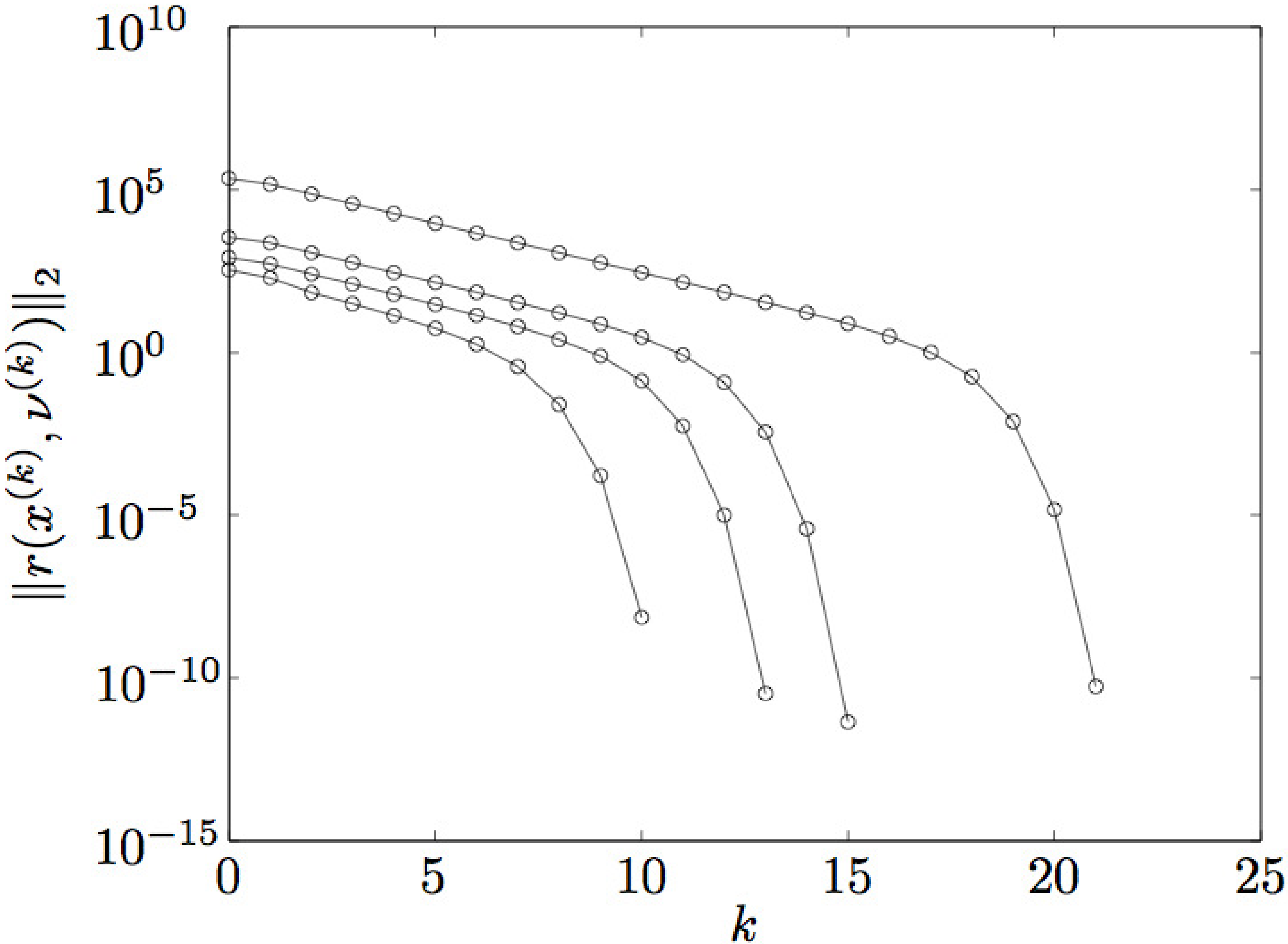
Example (cont)

1. Newton's method applied to the dual



Example (cont)

1. Infeasible start Newton's method



Summary

- The three methods have the *same complexity* for each iteration
- In this example, the *dual method* is faster, but only by a factor of 2 or 3.
- The methods also differ by the *initialization* they require:
  - Primal: $Ax^{(0)} = 0$, $x^{(0)} > 0$.
  - Dual: $A^\top \lambda^{(0)} > 0$.
  - Infeasible start: $x > 0$

  Depending on the problem, one or the other might be more readily available.