Definition 41 [Subgradient]: *Let $f : \mathcal{D} \rightarrow \Re$ be a convex function defined on a convex set $\mathcal{D}$. A vector $\mathbf{h} \in \Re^n$ is said to be a subgradient of f at the point $\mathbf{x} \in \mathcal{D}$ if*

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{h}^T(\mathbf{y} - \mathbf{x})$$

for all $\mathbf{y} \in \mathcal{D}$. The set of all such vectors is called the subdifferential of f at $\mathbf{x}$.

Theorem 76 *Let $f : \mathcal{D} \rightarrow \Re$ be a convex function defined on a convex set $\mathcal{D}$. A point $\mathbf{x} \in \mathcal{D}$ corresponds to a minimum if and only if*

$$\nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \geq 0$$

for all $\mathbf{y} \in \mathcal{D}$.

If $\nabla f(\mathbf{x})$ is nonzero, it defines a supporting hyperplane to $\mathcal{D}$ at the point $\mathbf{x}$. Theorem 77 implies that for a differentiable convex function defined on an open set, every critical point must be a point of (global) minimum.

Theorem 77 *Let $f : \mathcal{D} \rightarrow \Re$ be differentiable and convex on an open convex domain $\mathcal{D} \subseteq \Re^n$. Then $\mathbf{x}$ is a critical point of f if and only if it is a (global) minimum.*

Theorem 78 *Let $f : \mathcal{D} \rightarrow \Re$ with $\mathcal{D} \subseteq \Re^n$ be differentiable on the convex set $\mathcal{D}$. Then,*

1. *f is convex on $\mathcal{D}$ if and only if is its gradient ∇f is monotone. That is, for all $\mathbf{x}, \mathbf{y} \in \Re$*

$$\left(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\right)^T (\mathbf{x} - \mathbf{y}) \geq 0 \tag{4.53}$$

2. *f is strictly convex on $\mathcal{D}$ if and only if is its gradient ∇f is strictly monotone. That is, for all $\mathbf{x}, \mathbf{y} \in \Re$ with $\mathbf{x} \neq \mathbf{y}$,*

$$\left(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\right)^T (\mathbf{x} - \mathbf{y}) > 0 \tag{4.54}$$

3. *f is uniformly or strongly convex on $\mathcal{D}$ if and only if is its gradient ∇f is uniformly monotone. That is, for all $\mathbf{x}, \mathbf{y} \in \Re$,*

$$\left(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\right)^T (\mathbf{x} - \mathbf{y}) \geq c||\mathbf{x} - \mathbf{y}||^2 \tag{4.55}$$

for some constant $c > 0$.

Necessity: Suppose f is uniformly convex on $\mathcal{D}$. Then from theorem 75, we know that for any $\mathbf{x}, \mathbf{y} \in \mathcal{D}$,

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) - \frac{1}{2}c||\mathbf{y} + \mathbf{x}||^2$$
$$f(\mathbf{x}) \geq f(\mathbf{y}) + \nabla^T f(\mathbf{y})(\mathbf{x} - \mathbf{y}) - \frac{1}{2}c||\mathbf{x} + \mathbf{y}||^2$$

Adding the two inequalities, we get (4.55). If f is convex, the inequalities hold with $c = 0$, yielding (4.54). If f is strictly convex, the inequalities will be strict, yielding (4.54).

Sufficiency: Suppose ∇f is monotone. For any fixed $\mathbf{x}, \mathbf{y} \in \mathcal{D}$, consider the function $\phi(t) = f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))$. By the mean value theorem applied to $\phi(t)$, we should have for some $t \in (0, 1)$,

$$\phi(1) - \phi(0) = \phi'(t) \tag{4.56}$$

Letting $\mathbf{z} = \mathbf{x} + t(\mathbf{y} - \mathbf{x})$, (4.56) translates to

$$f(\mathbf{y}) - f(\mathbf{x}) = \nabla^T f(\mathbf{z})(\mathbf{y} - \mathbf{x}) \tag{4.57}$$

Also, by definition of monotonicity of ∇f, (from (4.53)),

$$(\nabla f(\mathbf{z}) - \nabla f(\mathbf{x}))^T (\mathbf{y} - \mathbf{x}) = \frac{1}{t} (\nabla f(\mathbf{z}) - \nabla f(\mathbf{x}))^T (\mathbf{z} - \mathbf{x}) \geq 0 \tag{4.58}$$

Combining (4.57) with (4.58), we get,

$$\begin{aligned} f(\mathbf{y}) - f(\mathbf{x}) &= (\nabla f(\mathbf{z}) - f(\mathbf{x}))^T (\mathbf{y} - \mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \\ &\geq \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \end{aligned} \tag{4.59}$$

By theorem 75, this inequality proves that f is convex. Strict convexity can be similarly proved by using the strict inequality in (4.58) inherited from strict monotonicity, and letting the strict inequality follow through to (4.59). For the case of strong convexity, from (4.55), we have

$$\begin{aligned} &\phi'(t) - \phi'(0) = (\nabla f(\mathbf{z}) - f(\mathbf{x}))^T (\mathbf{y} - \mathbf{x}) \\ &= \frac{1}{t} (\nabla f(\mathbf{z}) - f(\mathbf{x}))^T (\mathbf{z} - \mathbf{x}) \geq \frac{1}{t}c||\mathbf{z} - \mathbf{x}||^2 = ct||\mathbf{y} - \mathbf{x}||^2 \end{aligned} \tag{4.60}$$

$$\phi(1) - \phi(0) - \phi'(0) = \int_0^1 [\phi'(t) - \phi'(0)]dt \geq \frac{1}{2}c||\mathbf{y} - \mathbf{x}||^2 \tag{4.61}$$

which translates to

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \frac{1}{2}c||\mathbf{y} - \mathbf{x}||^2$$

$$\boxed{\begin{aligned}\min_x\ & f(x) \\ \text{s.t}\ & g_i(x) \le 0 \\ & h_j(x) = 0\end{aligned}} \equiv \boxed{\begin{aligned}\min_x\ & f(x) \\ \text{s.t}\ & g_i(x) \le 0 \\ & h_j(x) \le 0 \\ & -h_j(x) \le 0\end{aligned}}$$

$h_j(x)$ & $-h_j(x)$ are both convex $\Rightarrow h_j(x)$ is affine ie

$$[h(x)] = Ax + b = 0$$

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ b_p \end{bmatrix} \quad h_j(x) = a_j^T x + b_j$$

$$\begin{aligned}\min\ & f(x) \\ \text{st}\ & g_i(x) \le 0 \\ & Ax = b\end{aligned}$$

often $Ax \le b$

① inclusion
② exclusion

consider only "subset" of constraints (or "subsets" of variables) in each iteration

Active set methods ← contrast → **Interior point method**

Let x^k be st $g_{i_1}, g_{i_2} \ldots g_{i_k}(x^k) = 0$.
& assume $g_j(x^k) < 0 \ \forall$ other j

$$x^{k+1} = \underset{\text{st}\ g_{i_1}(x)=0\ g_{i_2}(x)=0 \ldots g_{i_k}(x)=0}{\operatorname{argmin}} f(x)$$

include / exclude constraints before next iter

Consider all inequality constraints but penalty associated with the constraints is included in the objective.

Equivalent convex problems

two problems are (informally) **equivalent** if the solution of one is readily obtained from the solution of the other, and vice-versa

some common transformations that preserve convexity:

- **eliminating equality constraints**

To solve analytically
OR
To apply descent methods

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

is equivalent to

$\nabla f_0(z) = F \nabla f_0(x)$
$\nabla f_i(z) = F \nabla f_i(x)$

$$\begin{array}{ll} \text{minimize (over } z\text{)} & f_0(Fz + x_0) \\ \text{subject to} & f_i(Fz + x_0) \le 0, \quad i = 1, \ldots, m \end{array}$$

where F and x_0 are such that

$$Ax = b \iff x = Fz + x_0 \text{ for some } z$$

Q: What if we want to invoke steepest descent with ∞ norm or 1 norm in $\|v\|_p = 1$?

Transformation of x to account for $Ax=b$

- **introducing equality constraints**

$$\begin{array}{ll} \text{minimize} & f_0(A_0 x + b_0) \\ \text{subject to} & f_i(A_i x + b_i) \le 0, \quad i = 1, \ldots, m \end{array}$$

is equivalent to

$$\begin{array}{ll} \text{minimize (over } x,\ y_i\text{)} & f_0(y_0) \\ \text{subject to} & f_i(y_i) \le 0, \quad i = 1, \ldots, m \\ & y_i = A_i x + b_i, \quad i = 0, 1, \ldots, m \end{array}$$

- **introducing slack variables for linear inequalities**

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & a_i^T x \le b_i, \quad i = 1, \ldots, m \end{array}$$

is equivalent to

$$\begin{array}{ll} \text{minimize (over } x,\ s\text{)} & f_0(x) \\ \text{subject to} & a_i^T x + s_i = b_i, \quad i = 1, \ldots, m \\ & s_i \ge 0, \quad i = 1, \ldots m \end{array}$$

Quadratic Optimization: Primal Active-Set Algorithm

Consider the quadratic optimization problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x} + \beta \\ \text{subject to} & A\mathbf{x} \geq \mathbf{b} \rightarrow a_i^T x \geq b_i \quad i=1..n \end{array} \qquad (1)$$

where $Q \succ 0$.

(A) I_0 = index set of active constraints, ie $\forall i \in I_0$ $a_i^T x^0 = b_i$ & $\forall$ $i \notin I_0$ $a_i^T x^0 > b_i$ (often initial $x^{(0)}$ is obtained using interior pt methods)

$k=0$

(B) $x^{(k+1)} = \underset{\text{st: } a_i^T x = b_i \quad i \in I_{(k)}}{\text{argmin}} \frac{1}{2}x^T Q x + c^T x + \beta \equiv \left(d^{(k+1)} = \underset{d}{\text{argmin}} \ \frac{1}{2}d^T Q d + g_k^T d \quad a_i^T d = 0 \quad i \in I_{(k)} \quad g_k = Qx^k + c \right)$

Substituting $x^{(k+1)} = x^{(k)} + d^{(k+1)}$

(C) If $x^{(k+1)}$ violates $a_i^T x^{(k+1)} \geq b_i$ for any $i \notin I_k$ then find α_k st $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k+1)}$ & $a_i^T x^{(k+1)} \geq b_i \ \forall i$

If we have overstepped

find $\alpha_k \in [0,1]$ to project back

(D) If convergence is not yet attained, update $I_{(k+1)}$ → should result in feasible soln for step (2)

Convergence in terms of KKT.. $\exists \lambda \geq 0$ st $\underbrace{Qx^{(k+1)} + c - A^T\lambda = 0}_{A^T\lambda = g^{(k+1)}}$ & $a_i x^{(k+1)} \geq b_i \ \forall i$

If $\alpha_k d^{(k+1)} = 0$ then it might need change in I_k to do better

If $\alpha_k < 1$ then some constraint has become active Add it to $I_{(k+1)}$

Might need to update $I_{(k+1)}$ even if $\alpha_k = 1$ since some constraint might have still become active.

Step 1
Input a feasible point, $\mathbf{x}^0$, identify the active set $\mathcal{I}^0$, form matrix $A_{\mathcal{I}^0}$, and set $k = 0$.

Step 2
Compute $\mathbf{g}^k = Q\mathbf{x}^k + \mathbf{c}$.
Check the rank condition $rank[A_{\mathcal{I}^k}^T \quad \mathbf{g}^k] = rank[A_{\mathcal{I}^k}^T]$. If it does not hold, go to **Step 4**.

Step 3
Convergence based on KKT condition in step D

Solve the system $A_{\mathcal{I}^k}^T \widehat{\lambda} = \mathbf{g}^k$. If $\widehat{\lambda} \geq \mathbf{0}$, output $\mathbf{x}^k$ as the solution and stop; otherwise, remove the index that is associated with the most negative Lagrange multiplier (some $\widehat{\lambda}_t$) from $\mathcal{I}^k$.

Remove a_t from active set (H/W: Justify)

Step 4
Compute the value of $\mathbf{d}^k$:

$$\mathbf{d}^k = \begin{array}{lll} \underset{\mathbf{d}}{\text{argmin}} & \frac{1}{2}\mathbf{d}^T Q \mathbf{d} + (\mathbf{g}^k)^T \mathbf{d} & \\ \text{subject to} & \mathbf{a}_i^T \mathbf{d} = 0 & \text{for } i \in \mathcal{I}^k \end{array} \tag{2}$$

Step 5
Compute α_k:

Projection of step C

$$\alpha_k = \min \left\{ 1, \quad \min_{\substack{j \notin \mathcal{I}^k \\ \mathbf{a}_j^T \mathbf{d}^k < 0}} \frac{\mathbf{a}_j^T \mathbf{x}^k - b_j}{-\mathbf{a}_j^T \mathbf{d}^k} \right\} \tag{3}$$

Set $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k \mathbf{d}^k$.

$\forall j \notin I_k \quad a_j^T(x^k + \alpha_k d^k) \geq b_j$
If $a_j^T d^k > 0$ ... no issues
If $a_j^T d^k < 0$
$\alpha_k (a_j^T d^k) \geq b_j - a_j^T x^k$
$\Rightarrow \alpha_k \leq \dfrac{a_j^T x^k - b_j}{(-a_j^T d^k)}$

Step 6
If $\alpha_k < 1$, construct $\mathcal{I}^{k+1}$ by adding the index that yields the minimum value of α_k in (3). Otherwise, let $\mathcal{I}^{k+1} = \mathcal{I}^k$.

Step 7
Set $k = k + 1$ and repeat from **Step 2**.

Figure 1: Optimization for the quadratic problem in (??) using Primal Active-set Method.

(contrasting with Active set)

12. Interior-point methods

- inequality constrained minimization
- logarithmic barrier function and central path
- barrier method
- feasibility and phase I methods

Inequality constrained minimization

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array} \qquad (1)$$

- f_i convex, twice continuously differentiable
- $A \in \mathbf{R}^{p \times n}$ with $\mathbf{rank}\, A = p$
- we assume $p^\star$ is finite and attained
- we assume problem is strictly feasible: there exists $\tilde{x}$ with

$$\tilde{x} \in \mathbf{dom}\, f_0, \qquad f_i(\tilde{x}) < 0, \quad i = 1, \ldots, m, \qquad A\tilde{x} = b$$

By slaters condition, strong duality

hence, strong duality holds and dual optimum is attained

Examples

- LP, QP, QCQP, GP
- entropy maximization with linear inequality constraints

$$
\begin{array}{ll}
\text{minimize} & \sum_{i=1}^{n} x_i \log x_i \\
\text{subject to} & Fx \preceq g \\
 & Ax = b
\end{array}
$$

Think of how to apply Active set to this problem

with $\mathbf{dom}\, f_0 = \mathbf{R}_{++}^n$

- differentiability may require reformulating the problem, *e.g.*, piecewise-linear minimization or ℓ_∞-norm approximation via LP
- SDPs and SOCPs are better handled as problems with generalized inequalities (see later)

Logarithmic barrier

reformulation of (1) via indicator function:

$$
\begin{array}{ll}
\text{minimize} & f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) \\
\text{subject to} & Ax = b
\end{array}
$$

where $I_-(u) = 0$ if $u \le 0$, $I_-(u) = \infty$ otherwise (indicator function of $\mathbf{R}_-$)

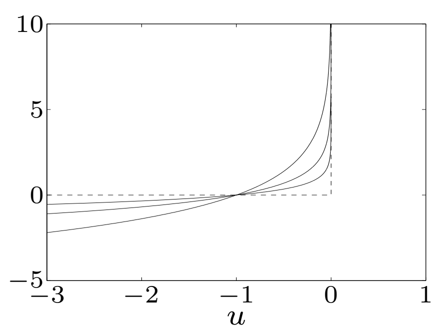
approximation via logarithmic barrier

$I_-(x) \approx -\log(-x)$

$I_-(f_i)$

$$
\begin{array}{ll}
\text{minimize} & f_0(x) - (1/t) \sum_{i=1}^{m} \log(-f_i(x)) \\
\text{subject to} & Ax = b
\end{array}
$$

- an equality constrained problem
- for $t > 0$, $-(1/t)\log(-u)$ is a smooth approximation of I_-
- approximation improves as $t \to \infty$



logarithmic barrier function

$$\phi(x) = -\sum_{i=1}^{m} \log(-f_i(x)), \qquad \mathbf{dom}\,\phi = \{x \mid f_1(x) < 0, \ldots, f_m(x) < 0\}$$

- convex (follows from composition rules)
- twice continuously differentiable, with derivatives

$$\begin{aligned}
\nabla \phi(x) &= \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x) \\
\nabla^2 \phi(x) &= \sum_{i=1}^{m} \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla^2 f_i(x)
\end{aligned}$$

Central path

- for $t > 0$, define $x^\star(t)$ as the solution of

$$\begin{array}{ll} \text{minimize} & t f_0(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

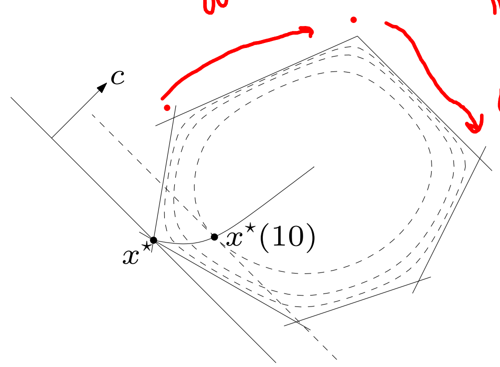
  (for now, assume $x^\star(t)$ exists and is unique for each $t > 0$)
- central path is $\{x^\star(t) \mid t > 0\}$

Diff between active set & interior pt

example: central path for an LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \le b_i, \quad i = 1, \ldots, 6 \end{array}$$

hyperplane $c^T x = c^T x^\star(t)$ is tangent to level curve of ϕ through $x^\star(t)$



Dual points on central path

$x = x^\star(t)$ if there exists a w such that

$$t\nabla f_0(x) + \sum_{i=1}^m \frac{1}{-f_i(x)}\nabla f_i(x) + A^T w = 0, \qquad Ax = b$$

- therefore, $x^\star(t)$ minimizes the Lagrangian

$$L(x, \lambda^\star(t), \nu^\star(t)) = f_0(x) + \sum_{i=1}^m \lambda_i^\star(t) f_i(x) + \nu^\star(t)^T(Ax - b)$$

  where we define $\lambda_i^\star(t) = 1/(-tf_i(x^\star(t))$ and $\nu^\star(t) = w/t$

- this confirms the intuitive idea that $f_0(x^\star(t)) \to p^\star$ if $t \to \infty$:

$$\begin{aligned} p^\star &\geq g(\lambda^\star(t), \nu^\star(t)) \\ &= L(x^\star(t), \lambda^\star(t), \nu^\star(t)) \\ &= f_0(x^\star(t)) - m/t \end{aligned}$$

Interpretation via KKT conditions

$x = x^\star(t)$, $\lambda = \lambda^\star(t)$, $\nu = \nu^\star(t)$ satisfy

1. primal constraints: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $Ax = b$
2. dual constraints: $\lambda \succeq 0$
3. approximate complementary slackness: $-\lambda_i f_i(x) = 1/t$, $i = 1, \ldots, m$
4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + A^T\nu = 0$$

difference with KKT is that condition 3 replaces $\lambda_i f_i(x) = 0$

H/W: How to find strictly feasible initial $x^{(0)}$?

Barrier method

Hint: Use interior pt

---

given strictly feasible x, $t := t^{(0)} > 0$, $\mu > 1$, tolerance $\epsilon > 0$.

repeat

1. *Centering step.* Compute $x^\star(t)$ by minimizing $tf_0 + \phi$, subject to $Ax = b$.
2. *Update.* $x := x^\star(t)$.
3. *Stopping criterion.* **quit** if $m/t < \epsilon$. upper bound on gap
4. *Increase* t. $t := \mu t$.

- terminates with $f_0(x) - p^\star \leq \epsilon$ (stopping criterion follows from $f_0(x^\star(t)) - p^\star \leq m/t$)
- centering usually done using Newton's method, starting at current x
- choice of μ involves a trade-off: large μ means fewer outer iterations, more inner (Newton) iterations; typical values: $\mu = 10$–20
- several heuristics for choice of $t^{(0)}$

Convergence analysis

number of outer (centering) iterations: exactly

$$\left\lceil \frac{\log(m/(\epsilon t^{(0)}))}{\log \mu} \right\rceil$$

plus the initial centering step (to compute $x^\star(t^{(0)})$)

centering problem

$$\text{minimize} \quad tf_0(x) + \phi(x)$$

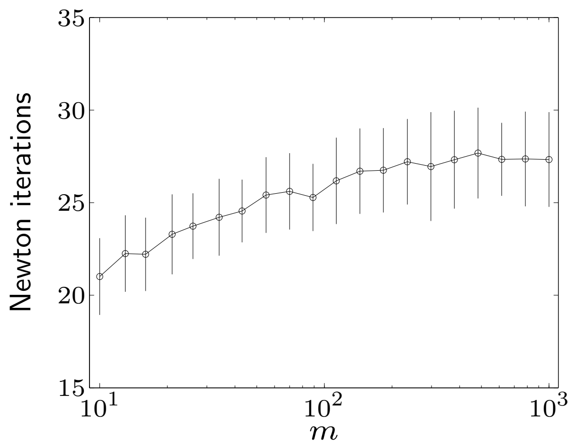
see convergence analysis of Newton's method

- $tf_0 + \phi$ must have closed sublevel sets for $t \geq t^{(0)}$
- classical analysis requires strong convexity, Lipschitz condition
- analysis via self-concordance requires self-concordance of $tf_0 + \phi$

family of standard LPs ($A \in \mathbf{R}^{m \times 2m}$)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \succeq 0 \end{array}$$

$m = 10, \ldots, 1000$; for each m, solve 100 randomly generated instances



number of iterations grows very slowly as m ranges over a $100 : 1$ ratio

Feasibility and phase I methods

feasibility problem: find x such that

$$f_i(x) \leq 0, \quad i = 1, \ldots, m, \qquad Ax = b \tag{2}$$

phase I: computes strictly feasible starting point for barrier method

basic phase I method

$$\begin{array}{ll} \text{minimize (over } x, s) & s \\ \text{subject to} & f_i(x) \leq s, \quad i = 1, \ldots, m \\ & Ax = b \end{array} \tag{3}$$

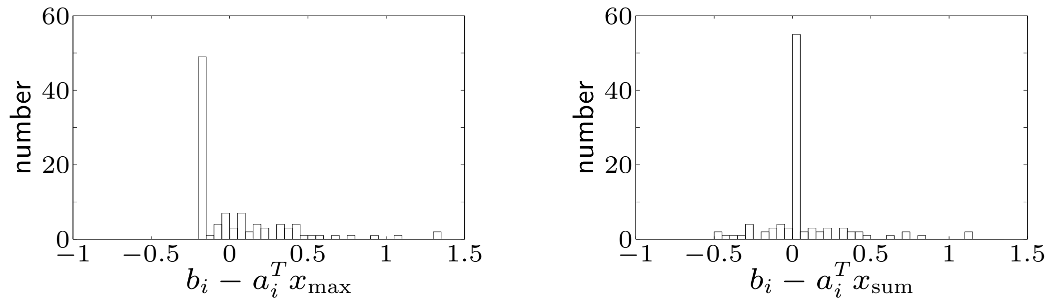
- if x, s feasible, with $s < 0$, then x is strictly feasible for (2)
- if optimal value $\bar{p}^\star$ of (3) is positive, then problem (2) is infeasible
- if $\bar{p}^\star = 0$ and attained, then problem (2) is feasible (but not strictly); if $\bar{p}^\star = 0$ and not attained, then problem (2) is infeasible

sum of infeasibilities phase I method

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T s \\ \text{subject to} & s \succeq 0, \quad f_i(x) \leq s_i, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

for infeasible problems, produces a solution that satisfies many more inequalities than basic phase I method

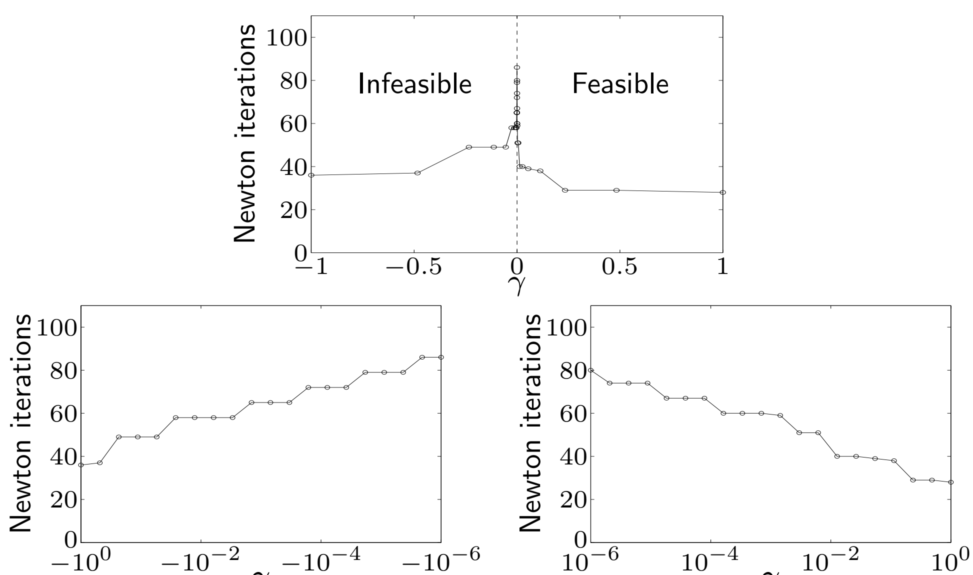
example (infeasible set of 100 linear inequalities in 50 variables)



left: basic phase I solution; satisfies 39 inequalities
right: sum of infeasibilities phase I solution; satisfies 79 solutions

example: family of linear inequalities $Ax \preceq b + \gamma \Delta b$

- data chosen to be strictly feasible for $\gamma > 0$, infeasible for $\gamma \leq 0$
- use basic phase I, terminate when $s < 0$ or dual objective is positive



number of iterations roughly proportional to $\log(1/|\gamma|)$