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

1. apply the active-set strategy, and explain why it can take 2^m linear solves;
2. write the barrier method, and specialize it to a quadratic objective with linear constraints;
3. show how decreasing μ increases the Newton decrement;
4. explain why μ can be divided by $1 + \frac{1}{20\sqrt{m}}$ after every Newton step.

1 Recap: Inequality constraints

Today's goal is to solve

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

Even computing local min is NP-hard. Practical algorithms seek **first-order optimality**, which reads

$$\nabla f_0(x)^T d \ge 0 \quad \text{for all } d \in \mathcal{T}(x), \tag{FOC}$$

where **tangent cone**

$$\mathcal{T}(x) = \{\dot\gamma(0) : \gamma(0) = 0, \quad \gamma : [0, 1) \to \Omega \text{ is feasible path}\}$$

is the set of initial velocities of feasible paths from x. Unfortunately, checking (FOC) is still coNP-complete in general.

Practical algorithms actually solve a **sufficient condition** for

(FOC), called the KKT equations

$$\text{(stationarity) } \nabla f_0(x) + \sum_{i=1}^{m} y_i \nabla f_i(x) = 0,$$
$$\text{(primal feasibility) } f_i(x) \leq 0 \text{ for all } i \in \{1, \ldots, m\},$$
$$\text{(dual feasibility) } y_i \geq 0 \text{ for all } i \in \{1, \ldots, m\},$$
$$\text{(complementarity) } y_i f_i(x) = 0 \text{ for all } i \in \{1, \ldots, m\}.$$

The last three equations imply two types of constraints

$$\text{inactive:} \quad f_i(x) < 0 \quad \text{and} \quad y_i = 0,$$
$$\text{active:} \quad f_i(x) = 0 \quad \text{and} \quad y_i \geq 0.$$

Once active constraints are known, solve in one shot by Lagrange multipliers with active constraints as equality constraints.

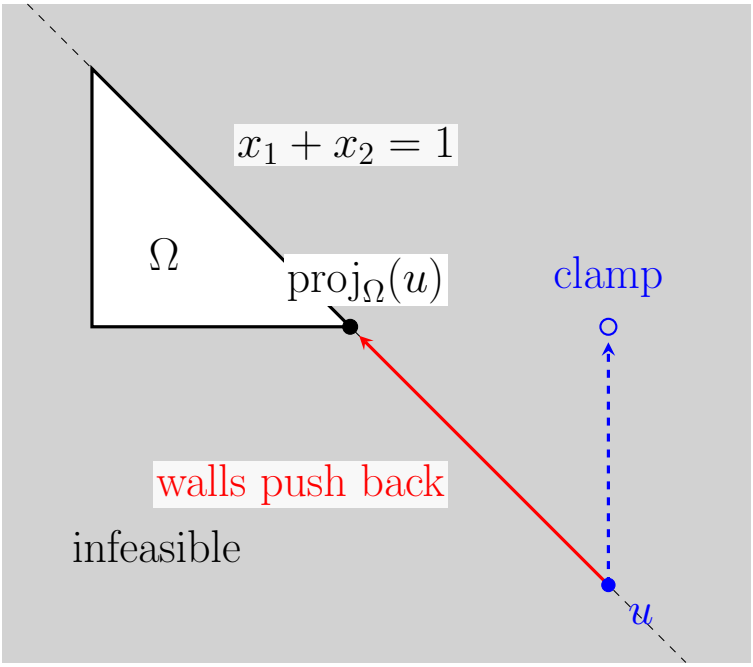
Active-set strategy: Try every choice of active set $\mathcal{A} \subseteq \{1, \ldots, m\}$.

- Set $y_i = 0$ for every $i \notin \mathcal{A}$, and $f_i(x) = 0$ for every $i \in \mathcal{A}$.
- Solve stationarity together with these equations, for x and for y_i with $i \in \mathcal{A}$.
- Check $y_i \geq 0$ for $i \in \mathcal{A}$, and $f_i(x) \leq 0$ for $i \notin \mathcal{A}$.

There are 2^m choices. For a quadratic objective and linear constraints, each choice is one linear system.

Recall the projection of $u = (2, -1)$ onto the triangle with walls

$$f_1(x) = -x_1, \qquad f_2(x) = -x_2, \qquad f_3(x) = x_1 + x_2 - 1.$$



Stationarity of $\frac{1}{2}\|x - u\|^2$ reads

$$x_1 - 2 - y_1 + y_3 = 0, \qquad x_2 + 1 - y_2 + y_3 = 0.$$

Every choice of active walls,

active walls	x	fails
none	$(2, -1)$	$f_2(x) = 1 > 0$
$\{1\}$	$(0, -1)$	$f_2(x) = 1 > 0$
$\{2\}$	$(2, 0)$	$f_3(x) = 1 > 0$
$\{3\}$	$(2, -1)$	$f_2(x) = 1 > 0$
$\{1, 2\}$	$(0, 0)$	$y_1 = -2 < 0$
$\{1, 3\}$	$(0, 1)$	$y_1 = -4 < 0$
$\{2, 3\}$	$(1, 0)$	nothing
$\{1, 2, 3\}$	no x	

Only $\{2, 3\}$ passes, so $\text{proj}_\Omega(u) = (1, 0)$ with $y = (0, 2, 1)$. With m walls, there are 2^m rows.

2 The barrier method

The barrier method replaces 2^m worst-case checks with guaranteed $O(\sqrt{m}\log\frac{1}{\epsilon})$ Newton steps, each costing $\mathrm{poly}(m)$ time. It solves the **barrier problem**

$$\min_{x\in\mathrm{Int}(\Omega)} F_\mu(x) = f_0(x) - \mu\sum_{i=1}^{m}\log\left(-f_i(x)\right),$$

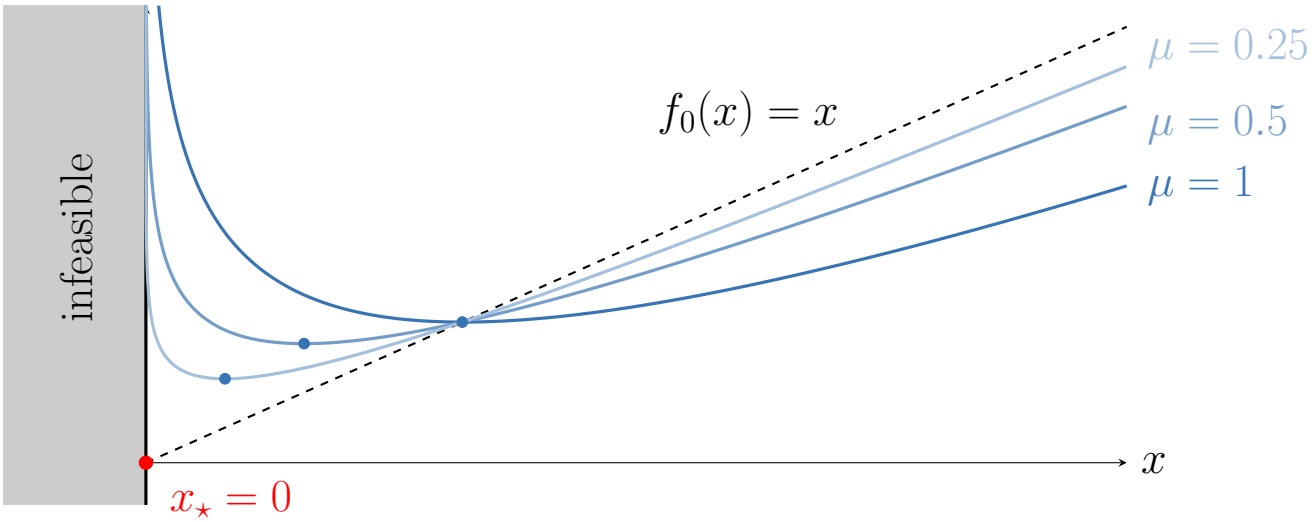
whose minimizer $x_\star(\mu)$, called the **center**, satisfies

$$f_0(x_\star(\mu)) - p^\star \le m\mu.$$

Given a strictly feasible $x \in \mathrm{Int}(\Omega)$, $\mu_0 > 0$, and $\rho > 1$, set $\mu = \mu_0$ and repeat

- **Centering.** Minimize F_μ by Newton's method, starting from the current x.
- **Stop** if $m\mu \le \epsilon$.
- **Update** $\mu \leftarrow \mu/\rho$.

For one variable, $\min_x x$ s.t. $-x \le 0$ has $F_\mu(x) = x - \mu\log x$.



3 Backtracking Newton

Backtracking Newton: Compute Newton search direction

$$v = -\nabla^2 f(x)^{-1} \nabla f(x),$$

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

$$\delta(x) := \sqrt{\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)}.$$

Lemma 1. Let f be L_1-Lipschitz gradient, L_2-Lipschitz Hessian, and β-strongly convex. Define

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

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

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

Problem: The log barrier is not L_1-Lipschitz gradient, not L_2-Lipschitz Hessian, and not β-strongly convex.

4 Self-concordant barriers

Quadratic objective, linear constraints: From now on, until the last section,

$$f_0(x) = \frac{1}{2}x^T Q x + c^T x \quad \text{with } Q \succeq 0, \qquad f_i(x) = a_i^T x - b_i.$$

It will be helpful to divide the barrier problem by μ,

$$F_\mu(x) = \frac{1}{\mu}\left(\frac{1}{2}x^T Q x + c^T x\right) - \sum_{i=1}^m \log\left(b_i - a_i^T x\right).$$

It has the same minimizer $x_\star(\mu)$, but it is **self-concordant**. Its Newton decrement is

$$\delta_\mu(x) = \sqrt{\nabla F_\mu(x)^T \nabla^2 F_\mu(x)^{-1} \nabla F_\mu(x)}.$$

Lemma 3. Let f be self-concordant with a minimizer $x_\star$. Define

$$\eta = \frac{1 - 2c_1}{4}, \qquad \gamma = \frac{c_1 c_2 \eta^2}{1+\eta}.$$

Backtracking Newton has two phases:
(Fixed decrement) If $\delta(x) > \eta$, then $f(x_+) \le f(x) - \gamma$.
(Quadratic convergence) Once $\delta(x) \le \eta$, then $\alpha = 1$ and $\delta(x_+) \le \frac{1}{2\eta}\delta(x)^2$.

Theorem 4. Under the same conditions as above, backtracking Newton converges to an iterate satisfying $\delta(x_k) \le \epsilon$ in at most

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

For the proofs, see Boyd and Vandenberghe, *Convex Optimization*, Section 9.6.4. The constants depend only on c_1 and c_2.

Question 1. Pick $c_1 = 0.1$ and $c_2 = \frac{1}{2}$. What is the iteration count for F_μ?

Solution: Substitute $c_1 = 0.1$ and $c_2 = \frac{1}{2}$,

$$\eta = \frac{1 - 2(0.1)}{4} = 0.2, \qquad \gamma = \frac{(0.1)(0.5)(0.2)^2}{1 + 0.2} = \frac{0.002}{1.2} = \frac{1}{600}.$$

Then $\frac{1}{\gamma} = 600$ and $2\eta = 0.4$, so $\delta_\mu(x_k) \le \epsilon$ within

$$\left\lceil 600\,(F_\mu(x_0) - F_\mu(x_\star(\mu))) + \log_2 \log_2 \frac{0.4}{\epsilon} \right\rceil \text{ iterations.}$$

5 Decreasing μ

Question 2. Why not pick a tiny μ and minimize F_μ directly?

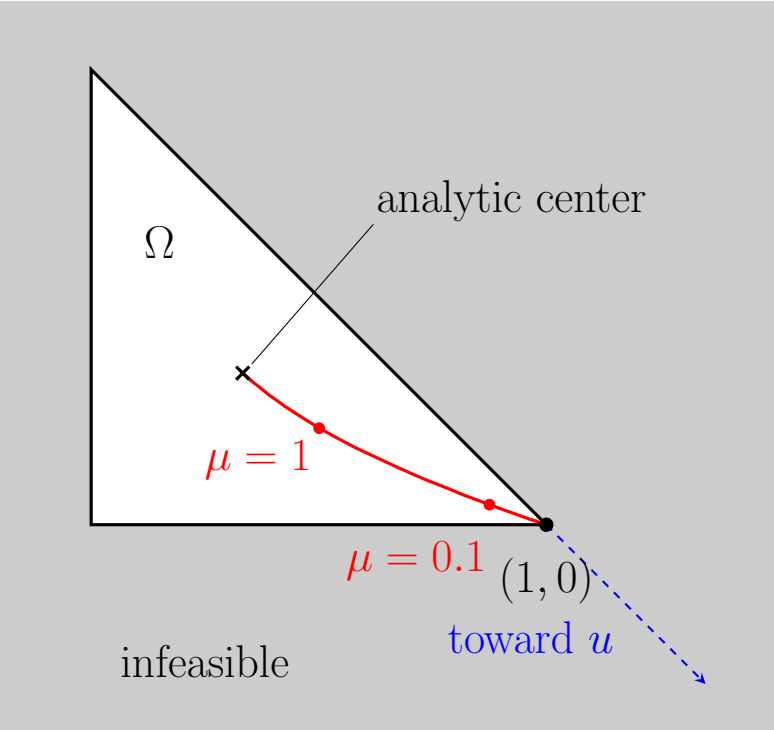
Solution: The count grows with $F_\mu(x_0) - F_\mu(x_\star(\mu))$, which is huge for tiny μ. For example, $\min_x x$ s.t. $-x \le 0$, divided by μ, gives

$$F_\mu(x) = \frac{x}{\mu} - \log x, \qquad F_\mu'(x) = \frac{1}{\mu} - \frac{1}{x} = 0 \quad \Longrightarrow \quad x_\star(\mu) = \mu.$$

From $x_0 = 1$,

$$F_\mu(1) = \frac{1}{\mu}, \qquad F_\mu(\mu) = 1 - \log \mu, \qquad F_\mu(1) - F_\mu(\mu) = \frac{1}{\mu} - 1 + \log \mu.$$

The gap grows like $\frac{1}{\mu}$. For $\mu = 10^{-6}$, it is about 10^6, and the bound is about 6×10^8 iterations.



Instead, start with a large μ, where F_μ is almost the barrier $-\sum_i \log(b_i - a_i^T x)$ alone. Then decrease μ a little, and take one Newton step. The centers $x_\star(\mu)$ form the **central path**.

The central path of the triangle projection runs from the **analytic center** $(\frac{1}{3}, \frac{1}{3})$, the minimizer of $-\sum_i \log(b_i - a_i^T x)$, to the solution $(1, 0)$.

Theorem 5. Decreasing μ. Let $Ax < b$ and $0 < \mu' \leq \mu$. Then

$$\delta_{\mu'}(x) \leq \frac{\mu}{\mu'}\,\delta_\mu(x) + \left(\frac{\mu}{\mu'} - 1\right)\sqrt{m}.$$

The proof is a homework problem.

Lemma 6. Quadratic convergence. Let $c_1 = 0.1$. If $\delta_\mu(x) \le 0.2$, then backtracking Newton on F_μ takes the full step $x_+ = x + v$, and

$$\delta_\mu(x_+) \le \frac{5}{2}\,\delta_\mu(x)^2.$$

Proof: With $c_1 = 0.1$, $\eta = 0.2$ and $\frac{1}{2\eta} = \frac{5}{2}$. This is the quadratic phase. □

Theorem 7. Short-step barrier method. Let $\delta_{\mu_0}(x_0) \le 0.2$. Iterate

$$x_{k+1} = x_k + v_k, \qquad \mu_{k+1} = \frac{\mu_k}{1 + \frac{1}{20\sqrt{m}}},$$

where v_k is the Newton step of F_{μ_k} at x_k. Then $\delta_{\mu_k}(x_k) \le 0.2$ for every k.

Proof: Let $\delta_{\mu_k}(x_k) \le 0.2$. By quadratic convergence,

$$\delta_{\mu_k}(x_{k+1}) \le \frac{5}{2}(0.2)^2 = 0.1.$$

Decreasing μ, with $\frac{\mu_k}{\mu_{k+1}} = 1 + \frac{1}{20\sqrt{m}}$,

$$\begin{aligned}\delta_{\mu_{k+1}}(x_{k+1}) &\le \left(1 + \frac{1}{20\sqrt{m}}\right) 0.1 + \frac{1}{20\sqrt{m}}\sqrt{m} \\ &\le 0.105 + 0.05 = 0.155 < 0.2.\end{aligned}$$

□

Each step is one linear solve, with no line search.

Theorem 8. Step count. In the short-step barrier method, $\mu_k \le \epsilon\mu_0$ after at most

$$\left\lceil (20\sqrt{m}+1)\log\frac{1}{\epsilon}\right\rceil \quad \text{Newton steps.}$$

Proof: After k steps, $\mu_k = \mu_0\rho^{-k}$ with $\rho = 1+\frac{1}{20\sqrt{m}}$. Take logarithms,

$$\log\frac{\mu_0}{\mu_k} = k\log\left(1+\frac{1}{20\sqrt{m}}\right).$$

Use $\log(1+t)\ge\frac{t}{1+t}$,

$$\log\left(1+\frac{1}{20\sqrt{m}}\right) \ge \frac{1/(20\sqrt{m})}{1+1/(20\sqrt{m})} = \frac{1}{20\sqrt{m}+1}.$$

So $\log\frac{\mu_0}{\mu_k}\ge\log\frac{1}{\epsilon}$ once $k\ge(20\sqrt{m}+1)\log\frac{1}{\epsilon}$. □

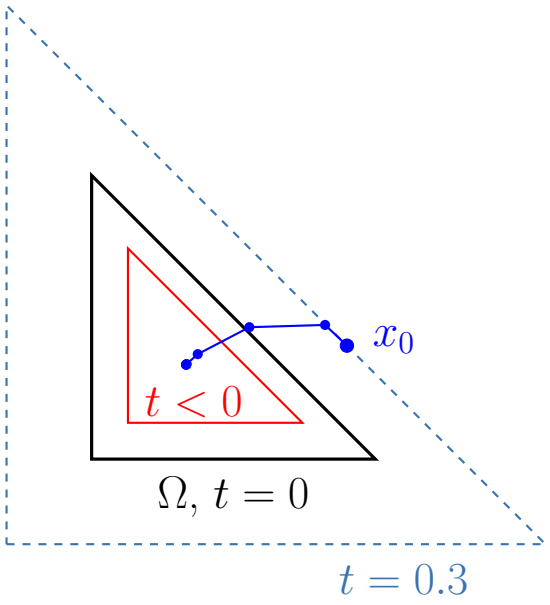
Points with $\delta_\mu(x)\le 0.2$ are within $2m\mu$ of optimal, stated without proof. So $f_0(x_k)-p^\star \le 2m\mu_0\epsilon$.

Pro: $O(\sqrt{m}\log\frac{1}{\epsilon})$ Newton steps, against 2^m active sets.

Con: For $m=10^6$, μ shrinks by only 0.005% per step.

6 Finding a starting point

The short-step method needs a strictly feasible x_0 with $\delta_{\mu_0}(x_0) \le 0.2$.



The start x_0 is outside Ω. Relax every constraint by t. The dashed triangle $\{x : a_i^T x - b_i \le 0.3 \text{ for all } i\}$ passes through x_0. Decreasing t shrinks the triangle, and the iterates (blue) follow. Once $t < 0$, the triangle (red) lies inside Ω, so x is strictly feasible. This is the **Phase I** problem

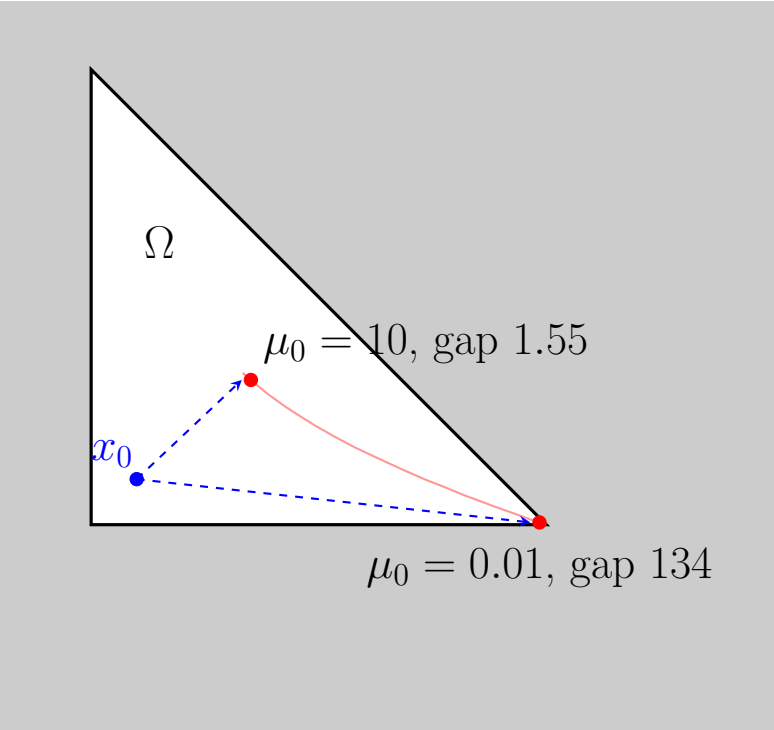
$$\min_{x,t} \ t \text{ s.t. } a_i^T x - b_i \le t \quad \text{for all } i.$$

Question 3. Find a strictly feasible point of the Phase I problem, starting from any x_0.

Solution: Take

$$t_0 = \max_i \left(a_i^T x_0 - b_i\right) + 1.$$

Then $a_i^T x_0 - b_i < t_0$ for every i. Run the barrier method on the Phase I problem from (x_0, t_0), and stop as soon as $t < 0$.



From $x_0 = (0.1, 0.1)$, the first centering moves toward $x_\star(\mu_0)$ on the central path (light red). Each label gives the gap $F_{\mu_0}(x_0) - F_{\mu_0}(x_\star(\mu_0))$.

Question 4. Why start with a large μ_0?

Solution: While $\delta_{\mu_0} > 0.2$, each backtracking step lowers F_{μ_0} by at least $\frac{1}{600}$, so $\delta_{\mu_0} \le 0.2$ after at most

$$600\,(F_{\mu_0}(x_0) - F_{\mu_0}(x_\star(\mu_0)))\ \text{steps.}$$

This is about $600 \times 134 \approx 80{,}000$ steps for $\mu_0 = 0.01$, and $600 \times 1.55 = 930$ steps for $\mu_0 = 10$. A large μ_0 costs only a few more short steps, $(20\sqrt{m} + 1)\log 10$ for each factor of 10.

7 Beyond quadratic objectives and linear constraints

For general f_0 and f_i, divide the barrier problem by μ as before,

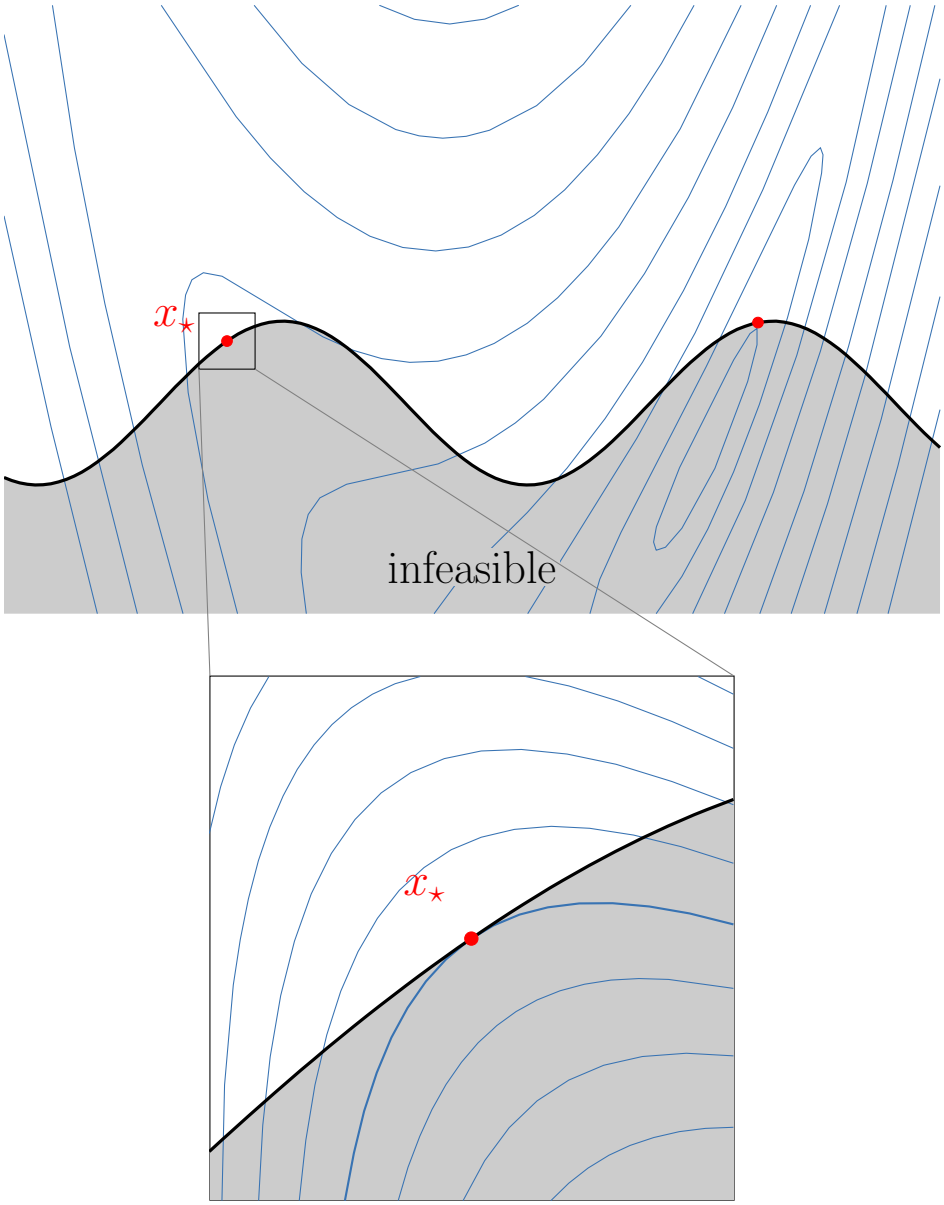
$$\min_x \; F_\mu(x) = \frac{f_0(x)}{\mu} - \sum_{i=1}^m \log\left(-f_i(x)\right).$$

Question 5. How can the quadratic-linear analysis be extended to nonconvex f_0 and f_i?

Solution: Zoom in on a local minimizer $x_\star$. Near $x_\star$, f_0 is close to its quadratic approximation, and each f_i is close to its linear approximation,

$$\begin{aligned} f_0(x) &\approx f_0(x_\star) + \nabla f_0(x_\star)^T(x - x_\star) \\ &\quad + \frac{1}{2}(x - x_\star)^T \nabla^2 f_0(x_\star)(x - x_\star), \\ f_i(x) &\approx f_i(x_\star) + \nabla f_i(x_\star)^T(x - x_\star). \end{aligned}$$

If the quadratic is convex, this is a quadratic objective with linear constraints. In a neighborhood where the approximations are accurate, Newton's method converges quadratically, and μ can be decreased by the same factor. We state this without proof. Outside such a neighborhood, there is no guarantee, and the method may approach a different local minimizer.



Above are the level curves of the nonconvex $f_0(x) = (1 - x_1)^2 + 2(x_2 - x_1^2)^2$, and a nonconvex feasible set $f_1(x) \leq 0$ in white. Its two local minimizers are red. Below, zoomed in on $x_\star$, the boundary $f_1 = 0$ 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.

8 Summary

The active-set strategy tries all 2^m choices of active constraints. The barrier method replaces the constraints by $-\mu \sum_i \log(-f_i(x))$, and its center is within $m\mu$ of optimal.

For a quadratic objective and linear constraints, dividing by μ gives F_μ, which is not strongly convex or Lipschitz but is self-concordant. Backtracking Newton then has the same two phases as for strongly convex functions. With $c_1 = 0.1$ and $c_2 = \frac{1}{2}$, its iteration bound is about $600(F_\mu(x_0) - F_\mu(x_\star(\mu)))$, which is huge for tiny μ.

Decreasing μ to μ' raises the decrement to at most $\frac{\mu}{\mu'}\delta + (\frac{\mu}{\mu'} - 1)\sqrt{m}$. With $\frac{\mu}{\mu'} = 1 + \frac{1}{20\sqrt{m}}$, one Newton step per decrease keeps $\delta_\mu \le 0.2$, and $(20\sqrt{m} + 1)\log\frac{1}{\epsilon}$ steps shrink μ by the factor ϵ.

Phase I finds a strictly feasible start, and a large μ_0 keeps the first centering short. For nonconvex f_0 and f_i, the analysis applies near a local minimizer, where the quadratic and linear approximations are accurate.