CS 2601 Linear and Convex Optimization

13. Barrier method

Bo Jiang

John Hopcroft Center for Computer Science
Shanghai Jiao Tong University

Fall 2022

General smooth convex problems

Consider

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{g}_j(\boldsymbol{x}) \leq 0, \ j = 1, 2, \ldots, m \end{aligned} \tag{ICP}$$

where f, g_j are smooth convex functions.

So far, we can solve some special cases

- When projection onto the feasible set is easy, we can use projected gradient descent
- When there are no inequality constraints, we can use Newton's method or projected gradient descent

For the general (ICP), we are going to get rid of the inequality constraints, but account for their effects by modifying the objective function.

Reformulation

Let $\bar{S}$ be the set defined by the inequality constraints,

$$\bar{S} = \{\boldsymbol{x} : g_j(\boldsymbol{x}) \leq 0, j = 1, 2, \ldots, m\}$$

(ICP) is equivalent to

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) + I_{\bar{S}}(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

where $I_{\bar{S}}(\boldsymbol{x})$ is the indicator of $\bar{S}$

$$I_{\bar{S}} = \begin{cases} 0, & \boldsymbol{x} \in \bar{S} \\ +\infty, & \boldsymbol{x} \notin \bar{S} \end{cases}$$

Interior-point methods approximate $I_{\bar{S}}(\boldsymbol{x})$ by so-called barrier functions.

Barrier function

Let S be the interior of $\bar{S}$,

$$S = \{\boldsymbol{x} : g_j(\boldsymbol{x}) < 0, j = 1, 2, \ldots, m\}$$

Assume it is nonempty. A continuous function $B : S \to \mathbb{R}$ is a barrier function for S if $B(\boldsymbol{x}) \to +\infty$ as $\boldsymbol{x} \in S$ approaches the boundary of S

Common barrier functions

- logarithmic barrier function

$$B(\boldsymbol{x}) = -\sum_{j=1}^{m} \log(-g_j(\boldsymbol{x}))$$

- inverse barrier function

$$B(\boldsymbol{x}) = -\sum_{j=1}^{m} \frac{1}{g_j(\boldsymbol{x})}$$

Note both barrier functions are convex.

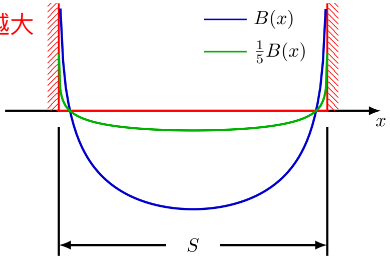
Log barrier function

Example. $S = (a, b)$ with

$$g_1(x) = a - x < 0$$
$$g_2(x) = x - b < 0$$

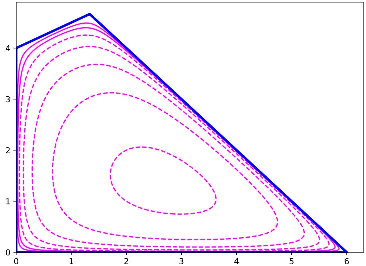
$$B(x) = -\log(x-a) - \log(b-x)$$

1/t B(x) t越大越接近 indicator function



Example. $S = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} < \boldsymbol{b}\}$, where

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ -1 & 2 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 6 \\ 8 \\ 0 \\ 0 \end{bmatrix}$$



$$B(\boldsymbol{x}) = -\log(6 - x_1 - x_2) - \log(8 + x_1 - 2x_2) - \log(x_1) - \log(x_2)$$

Log barrier function (cont'd)

$$B(\boldsymbol{x}) = -\sum_{j=1}^{m} \log(-g_j(\boldsymbol{x}))$$

gradient

$$\nabla B(\boldsymbol{x}) = \sum_{j=1}^{m} \frac{1}{-g_j(\boldsymbol{x})} \nabla g_j(\boldsymbol{x})$$

Hessian

$$\nabla^2 B(\boldsymbol{x}) = \sum_{j=1}^{m} \frac{1}{g_j^2(\boldsymbol{x})} \nabla g_j(\boldsymbol{x}) \nabla g_j(\boldsymbol{x})^T + \sum_{j=1}^{m} \frac{1}{-g_j(\boldsymbol{x})} \nabla^2 g_j(\boldsymbol{x})$$

Proof. Suffices to prove for $m = 1$ i.e. $B(\boldsymbol{x}) = -\log(-g(\boldsymbol{x}))$. Use

$$\partial_\ell \left(\frac{\partial_k g}{-g} \right) = \frac{\partial_\ell g \partial_k g}{g^2} + \frac{\partial_\ell \partial_k g}{-g} = \frac{[\nabla g \nabla g^T]_{k\ell}}{g^2} + \frac{[\nabla^2 g]_{k\ell}}{-g}$$

Interior-point methods

Interior-point methods approximate $I_{\bar{S}}(\boldsymbol{x})$ by $\frac{1}{t}B(\boldsymbol{x})$ and (ICP) by

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & F(\boldsymbol{x}) = f(\boldsymbol{x}) + \frac{1}{t}B(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned} \tag{B}$$

where the parameter $t > 0$ controls the approximation accuracy.

Note. For (B) to be feasible, (ICP) should be strictly feasible i.e. $\{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, g_j(\boldsymbol{x}) < 0, j = 1, \ldots, m\}$ is nonempty. We assume so.

We can solve (B) by e.g. damped Newton's method.

Note. $\operatorname{dom} F \neq \mathbb{R}^n$ (why?). Always check whether $\boldsymbol{x}_k + t\boldsymbol{d}_k \in \operatorname{dom} F$ during backtracking line search. You can either modify line 5 of §12, slide 12, or define $F(\boldsymbol{x}) = +\infty$ for $\boldsymbol{x} \notin \operatorname{dom} F$.

Example: LP

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^2} \quad & f(\boldsymbol{x}) = -x_1 - 3x_2 \\ \text{s.t.} \quad & x_1 + x_2 \leq 6 \\ & -x_1 + 2x_2 \leq 8 \\ & x_1, x_2 \geq 0 \end{aligned}$$

The approximation is an unconstrained problem

$$\min \; F(\boldsymbol{x})$$

with

$$F(\boldsymbol{x}) = -x_1 - 3x_2 + \frac{1}{t}[-\log(6 - x_1 - x_2) - \log(8 + x_1 - 2x_2) - \log(x_1) - \log(x_2)]$$

which we can solve by gradient descent or Newton's method.

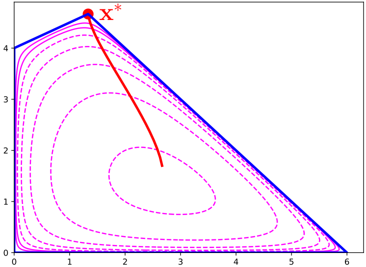
Note. We actually have implicit constraints.

$$\operatorname{dom} F = S = \{\boldsymbol{x} : x_1 + x_2 < 6, -x_1 + 2x_2 < 8, x_1 < 0, x_2 < 0\} \neq \mathbb{R}^2$$

Central path

The solution $\boldsymbol{x}^*(t)$ to (B) is called a central point. As $t > 0$ varies, the curve $\boldsymbol{x}^*(t)$ is called the central path of (ICP).

粉色的是
barrier 的level
curves

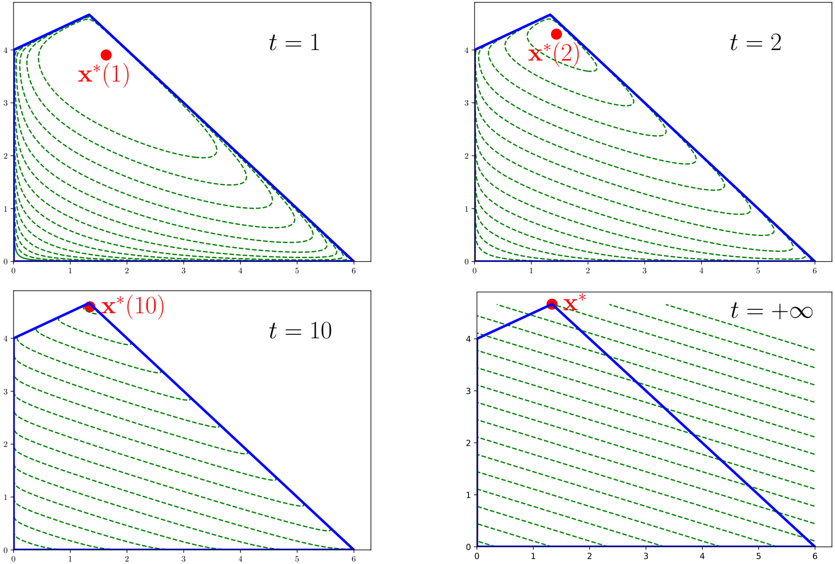


Note. $\boldsymbol{x}^*(t) \in S$ is an interior point of the set defined by the inequality constraints, whence the name interior-point method.

Note. $\boldsymbol{x}^*(t)$ is always a strictly feasible solution to (ICP). As $t \to \infty$, $\boldsymbol{x}^*(t)$ converges to the optimal solution $\boldsymbol{x}^*$ of (ICP).

Central point

Central points and level curves of $f(\boldsymbol{x}) + \frac{1}{t}B(\boldsymbol{x})$ for various values of t.



Suboptimality of central point

Lemma. The solution $\boldsymbol{x}^*(t)$ to (B) is $\frac{m}{t}$-suboptimal for (ICP), i.e.

$$f(\boldsymbol{x}^*(t)) - f^* \leq \frac{m}{t}$$

Proof. By the Lagrange condition for the problem (B), there exists $\boldsymbol{\lambda}^*$ s.t.

$$\nabla f(\boldsymbol{x}^*(t)) + \frac{1}{t}\nabla B(\boldsymbol{x}^*(t)) + \boldsymbol{A}^T\boldsymbol{\lambda}^* = \boldsymbol{0}$$

Using the formula for ∇B on slide 4, we obtain

$$\nabla f(\boldsymbol{x}^*(t)) + \sum_{j=1}^{m} \mu_j^* \nabla g_j(\boldsymbol{x}^*(t)) + \boldsymbol{A}^T\boldsymbol{\lambda}^* = \boldsymbol{0} \qquad (\dagger)$$

where $\mu_j^* = -\frac{1}{t g_j(\boldsymbol{x}^*(t))}$. Note $\mu_j^* > 0$, since $\boldsymbol{x}^*(t)$ is strictly feasible.

Let $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ be the Lagrangian of (ICP). Then $\nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \boldsymbol{0}$ by (†). Thus $\boldsymbol{x}^*(t) = \operatorname{argmin}_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, since $\mathcal{L}$ is convex in $\boldsymbol{x}$.

Suboptimality of central point (cont'd)

For $\boldsymbol{\mu} \geq \mathbf{0}$ and $\boldsymbol{x}$ in the feasible set X,

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{j=1}^{m} \underbrace{\mu_j}_{\geq 0} \underbrace{g_j(\boldsymbol{x})}_{\leq 0} + (\underbrace{\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}}_{=\mathbf{0}})^T \boldsymbol{\lambda} \leq f(\boldsymbol{x}) \qquad (\ddagger)$$

In particular,

$$f(\boldsymbol{x}) \geq \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*), \quad \forall \boldsymbol{x} \in X$$

Minimizing over feasible $\boldsymbol{x}$,

$$\begin{aligned}
f^* = \inf_{\boldsymbol{x} \in X} f(\boldsymbol{x}) &\geq \inf_{\boldsymbol{x} \in X} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \quad \text{by } (\ddagger) \\
&\geq \mathcal{L}(\boldsymbol{x}^*(t), \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \quad \text{since } \boldsymbol{x}^*(t) \text{ minimizes } \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \\
&= f(\boldsymbol{x}^*(t)) + \sum_{j=1}^{m} \underbrace{\mu_j^* g_j(\boldsymbol{x}^*(t))}_{=1/t \text{ by def. of } \mu_j^*} + (\underbrace{\boldsymbol{A}\boldsymbol{x}^*(t) - \boldsymbol{b}}_{=\mathbf{0} \text{ by feasibility}})^T \boldsymbol{\lambda}^* \\
&= f(\boldsymbol{x}^*(t)) - \frac{m}{t}
\end{aligned}$$

Interpretation via modified KKT conditions

$\boldsymbol{x}^*(t)$ satisfies the following modified KKT conditions for (ICP),

- feasibility

$$\boldsymbol{A}\boldsymbol{x}^*(t) = \boldsymbol{b}, \quad g_j(\boldsymbol{x}^*(t)) < 0,\ j = 1, 2, \ldots, m$$

- nonnegativity

$$\mu_j^* > 0, \quad j = 1, 2, \ldots, m$$

- stationarity

$$\nabla f(\boldsymbol{x}^*(t)) + \boldsymbol{A}^T \boldsymbol{\lambda}^* + \sum_{j=1}^{m} \mu_j^* \nabla g_j(\boldsymbol{x}^*(t)) = \boldsymbol{0}$$

- approximate complementary slackness

$$\mu_j^* g_j(\boldsymbol{x}^*(t)) = -\frac{1}{t} \approx 0 \text{ for large } t, \quad j = 1, 2, \ldots, m$$

Barrier method

Since $\boldsymbol{x}^*(t)$ is $\frac{m}{t}$-suboptimal, to achieve a desired accuracy, it suffices to solve

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) + \frac{\epsilon}{m} B(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

However, this works well only for small problems, good initial points, and moderate accuracy (ϵ not too large).

The barrier method or path-following method chooses an increasing sequence $\{t_k\}$ and solves a sequence of problems, until $t_k \geq m/\epsilon$,

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) + \frac{1}{t_k} B(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

When solving for the k-th central point $\boldsymbol{x}^*(t_k)$, it uses the central point $\boldsymbol{x}^*(t_{k-1})$ as the initial point.

Barrier method (cont'd)

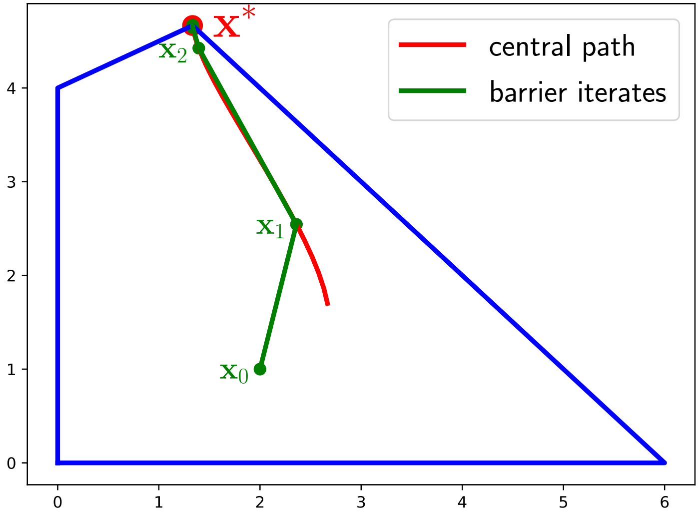
1: Choose a strictly feasible $\boldsymbol{x}_0$ s.t. $\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{b}$ and $\boldsymbol{g}(\boldsymbol{x}_0) < \boldsymbol{0}$, and $t_0 > 0$
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: $\quad \boldsymbol{x}_{k+1} \leftarrow \underset{\boldsymbol{x}:\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}}{\operatorname{argmin}}\{f(\boldsymbol{x}) + \frac{1}{t_k}B(\boldsymbol{x})\}$ starting from $\boldsymbol{x}_k$
4: $\quad$ **if** $t_k \geq m/\epsilon$ **then**
5: $\quad\quad$ **return** $\boldsymbol{x}_{k+1}$
6: $\quad$ **end if**
7: $\quad t_{k+1} \leftarrow \rho t_k$
8: **end for**

where $\rho > 1$ and B is the log barrier function.

- On Line 3, we can solve the equality constrained problem by Newton's method or projected gradient descent. Newton's method is often recommended to better deal with ill-conditioning.
- The choice of ρ and t_0 may affect how fast the algorithm converges. In practice, the performance is quite robust to the choice of ρ and t_0. [BV] suggests $\rho \in [10, 20]$

Example: LP

Solve the LP on slide 7 with $t_0 = 0.3$, $\boldsymbol{x}_0 = (2,1)^T$, $\rho = 10$.



Feasibility problem

How to find a strictly feasible initial point? By solving

$$\begin{aligned} \min_{\boldsymbol{x},s} \quad & s \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & g_j(\boldsymbol{x}) \le s, \; j = 1, 2, \ldots, m \end{aligned} \tag{F}$$

The original (ICP) is strictly feasible iff the optimal value s^* of (F) is negative.

How do we solve (F)? By the barrier method, since it is easy to to find a strictly feasible point for (F).

- Pick[1] an $\boldsymbol{x}_0$ satisfying $\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{b}$ and then pick an $s_0 > \max_{1 \le j \le m} g_j(\boldsymbol{x}_0)$.

We do not need to solve (F) until optimality. Once we find a feasible (x, s) with $s < 0$, we can stop and use $\boldsymbol{x}$ as the initial point for (ICP).

[1] assume $\operatorname{dom} g_j = \mathbb{R}^n$, $\forall j$; otherwise, we need a little bit extra work.