Interior Point Method

Convex Optimization, Lecture 14

by Sergei Savin

Spring 2025

Content

- Newton method
- Barrier functions
- Interior point method
- Analytic center of linear inequalities

Newton method, 1

Given a function $f : \mathbb{R} \to \mathbb{R}$, we can write it as $y = f(x)$. Consider the problem of finding roots of the equation $f(x) = 0$.

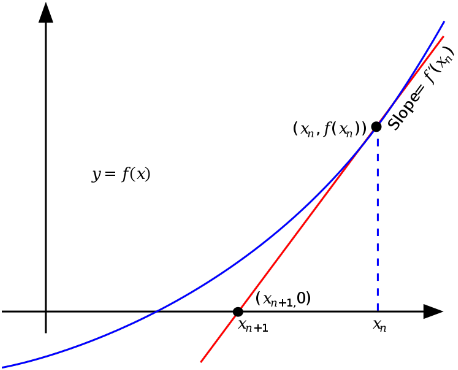
To solve it we can use an iterative procedure. We start at the initial guess x_0, produce affine approximation of the function at this point using constant and linear terms of its Taylor expansion:

$$f = f(x_0) + \frac{\partial f}{\partial x}(x - x_0) + \text{h.o.t.}, \tag{1}$$

and then solve the problem $f(x_0) + \frac{\partial f}{\partial x}(x - x_0) = 0$, to find a new point x_1, which will serve the role of x_0 during the next iteration. This is called *Newton's method*.

The solution to the equation $f(x_0) + \frac{\partial f}{\partial x}(x - x_0) = 0$ is:

$$x = x_0 - \left(\frac{\partial f}{\partial x}\right)^{-1} f(x_0) \tag{2}$$



Given a function $\varphi : \mathbb{R} \to \mathbb{R}$, we can write it as $y = \varphi(x)$. Let us find minimum of this function.

We can try to do it by looking at its quadratic approximation (considering constant, linear and quadratic terms of the Taylor expansion) at an approximation point x_0:

$$\varphi = \varphi(x_0) + \frac{\partial \varphi}{\partial x}(x - x_0) + \frac{1}{2}(x - x_0)\frac{\partial^2 \varphi}{\partial x^2}(x - x_0) + \text{h.o.t.}, \quad (3)$$

and finding its minimum, which can be done by solving a single least-squares problem. Its solution x_1 becomes the next approximation point.

This is *another form of the Newton's method*, and is completely equivalent to the previously discussed one.

Let us define $g = \frac{\partial \varphi}{\partial x}\Big|_{x=x_0}$ and $H = \frac{\partial^2 \varphi}{\partial x^2}\Big|_{x=x_0}$. Let us solve the problem:

$$\text{minimize:} \quad \varphi(x_0) + g(x - x_0) + \frac{1}{2}(x - x_0)H(x - x_0) \tag{4}$$

The minimum is attained when the derivative is achieves zero:

$$g + (x - x_0)H = 0 \tag{5}$$

This gives us solution $x = x_0 - H^{-1}g$.

Let us define a function $p(x) = \frac{\partial \varphi}{\partial x}$. Function $\varphi(x)$ achieves minimum when $p(x)$ crosses zero. Solving the problem $p(x) = 0$ with Newton's method, we initiate iterative process:

$$x_{i+1} = x_i - \left(\frac{\partial p}{\partial x} \right)^{-1} p(x_i) \tag{6}$$

Given the definition of $p(x)$ we can re-write this expression as:

$$x_{i+1} = x_i - \left(\frac{\partial^2 \varphi}{\partial x^2} \right)^{-1} \frac{\partial \varphi}{\partial x} \tag{7}$$

Using previously defined variables $g = \frac{\partial \varphi}{\partial x}$ and $H = \frac{\partial^2 \varphi}{\partial x^2}$ we obtain the same expression we saw in the sequential quadratic approximation:

$$x_{i+1} = x_i - H^{-1} g \tag{8}$$

For mutivaribe case $\varphi : \mathbb{R}^n \to \mathbb{R}$ the Taylor expansion is given exactly the same:

$$\varphi = \varphi(\mathbf{x}_0) + \frac{\partial \varphi}{\partial \mathbf{x}}(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)\frac{\partial^2 \varphi}{\partial \mathbf{x}^2}(\mathbf{x} - \mathbf{x}_0) + \text{h.o.t.}, \quad (9)$$

Solution to the quadratic approximation is given as:

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \mathbf{H}^{-1}\mathbf{g} \quad (10)$$

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \left(\frac{\partial^2 \varphi}{\partial \mathbf{x}^2}\right)^{-1} \frac{\partial \varphi}{\partial \mathbf{x}} \quad (11)$$

A key feature of Neuton's method is very fast convergence near the solution.

Solving constrained optimization

Newton's method works for *unconstrained optimization*.

Linear equalities in convex programs can often be excluded by introducing new variables.

Inequality constraints can be replaces with *soft constraints* - an addition to the cost function which punishes decision variables that approach the border of the domain.

Consider linear inequality constraints:

$$\mathbf{A}\mathbf{x} \leq \mathbf{b} \tag{12}$$

Remember that we can rewrite it as:

$$\mathbf{a}_i^\top \mathbf{x} \leq b_i \tag{13}$$

$$\mathbf{a}_i^\top \mathbf{x} - b_i \leq 0 \tag{14}$$

Instead of *hard constraints* in (14) we can turn these into a cost function component:

$$J = -\sum_{i=1}^{n} \log(b_i - \mathbf{a}_i^\top \mathbf{x}) \tag{15}$$

Which is called a *barrier function*.

Let us consider barrier functions $J = -\sum_{i=1}^{n} \log(b_i - \mathbf{a}_i^\top \mathbf{x})$:

- It removes the constraint, but modifies the cost.
- When $b_i - \mathbf{a}_i^\top \mathbf{x}$ is a very small positive number, $\log(b_i - \mathbf{a}_i^\top \mathbf{x})$ is a very big negative number, hence the minus sign in front.
- Barrier function does not behave well outside of the domain, when $b_i - \mathbf{a}_i^\top \mathbf{x} < 0$.

Hence the following QP:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{x}^\top \mathbf{H} \mathbf{x} + \mathbf{f}^\top \mathbf{x}, \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \end{aligned} \tag{16}$$

...can be approximated as:

$$\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{x}^\top \mathbf{H} \mathbf{x} + \mathbf{f}^\top \mathbf{x} - \sum_{i=1}^{n} \log(b_i - \mathbf{a}_i^\top \mathbf{x}) \tag{17}$$

Zero-infinity function can be seen as a precise barrier - adding nothing to the cost in the domain interior and punishing constraint violation by infinite increase of cost:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \infty & x > 0 \end{cases} \tag{18}$$

This can be written as max over a family of linear functions:

$$f(x) = \max_{\lambda \geq 0} \lambda x \tag{19}$$

The QP discussed in the previous example becomes:

$$\underset{\mathbf{x}}{\text{minimize}} \max_{\lambda \geq 0} \quad \mathbf{x}^\top \mathbf{H} \mathbf{x} + \mathbf{f}^\top \mathbf{x} - \sum_{i=1}^{n} \lambda_i (b_i - \mathbf{a}_i^\top \mathbf{x}) \tag{20}$$

It appears that we re-discovered Lagrangian method.

Interior point method is based on KKT conditions with a single modification highlifgted below:

1. Lagrangian stationarity: $\frac{\partial f_0(\mathbf{x})}{\partial \mathbf{x}} + \sum_i \lambda_i \frac{\partial g_i(\mathbf{x})}{\partial \mathbf{x}} = 0$.
2. Primal feasibility: $g_i(\mathbf{x}) \leq 0$.
3. Dual feasibility: $\lambda_i \geq 0$.
4. Complementarity slackness: $\lambda_i g_i(\mathbf{x}) = -\mathbf{t}_i$.

From the complementarity we can find expression for lagrange multipliers:

$$\lambda_i = -\frac{t_i}{g_i(\mathbf{x})} \tag{21}$$

Given $\lambda_i = -\frac{t_i}{g_i(\mathbf{x})}$ we can transform the Lagrangian stationarity condition:

$$\frac{\partial f_0(\mathbf{x})}{\partial \mathbf{x}} - \sum_i \frac{t_i}{g_i(\mathbf{x})} \frac{\partial g_i(\mathbf{x})}{\partial \mathbf{x}} = 0 \tag{22}$$

Let us compute the following derivative:

$$\frac{\partial}{\partial \mathbf{x}}(-t \log(-g(\mathbf{x}))) = \frac{-t}{-g(\mathbf{x})}(-1)\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} = -\frac{t}{g(\mathbf{x})}\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} \tag{23}$$

With that, we can re-write the lagrangain condition:

$$\frac{\partial f_0(\mathbf{x})}{\partial \mathbf{x}} - \frac{\partial}{\partial \mathbf{x}} \left(\sum_i t_i \log(-g_i(\mathbf{x})) \right) = 0 \tag{24}$$

Condition $\frac{\partial f_0(\mathbf{x})}{\partial \mathbf{x}} - \frac{\partial}{\partial \mathbf{x}} \left(\sum_i t_i \log(-g_i(\mathbf{x})) \right) = 0$ is equivalent to minimization:

$$\underset{\mathbf{x}}{\text{minimize}} \quad f_0(\mathbf{x}) - \sum_i t_i \log(-g_i(\mathbf{x})) \tag{25}$$

Solution to this problem $\mathbf{x}^*$ implies that $\log(-g_i(\mathbf{x}))$ exists, meaning that $g_i(\mathbf{x}) \leq 0$, giving us primal feasibility.

In the same way, $\lambda_i = -\frac{t_i}{g_i(\mathbf{x})} \geq 0$, giving us dual feasibility.

Thus, we have all KKT conditions satisfied automatically by solving the minimization (25).

Interior point method consists of constructing problem (25) and solving it using Newton's method for a given value of t_i; the method is applied iteratively, with dwindling t_i over subsequent iterations. Solution $\mathbf{x}^*$ of the previous iteration is used to initialize Newton's method on the next iteration.

The sequence of $\mathbf{x}^*$ resulting from this iterative process is called *central path.*

As t_i approaches zero, the log barriers $t_i \log(-g_i(\mathbf{x}))$ show less and less influence on the position of the optimal solution, allowing us to approximate the solution to the original problem more and more accurately.

We can define *analytic center of linear inequalities* as a minimum of the function $J = -\sum_{i=1}^{n} \log(b_i - \mathbf{a}_i^\top \mathbf{x})$. And that can be solved as a convex optimization:

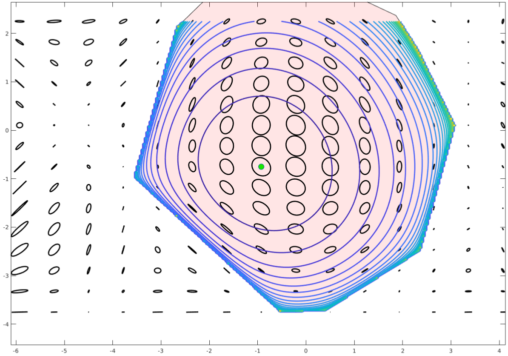
$$\mathbf{x}_a = \underset{\mathbf{x}}{\operatorname{argmin}} \quad - \sum_{i=1}^{n} \log(b_i - \mathbf{a}_i^\top \mathbf{x})$$

At the analytic center of linear inequalities the shape of contour lines can be analysed as a local quadratic approximation of the function J:

$$\mathcal{C} = \{\mathbf{x} : \; (\mathbf{x} - \mathbf{x}_a)^\top \frac{\partial^2 J}{\partial \mathbf{x}^2} (\mathbf{x} - \mathbf{x}_a) = \epsilon\} \tag{26}$$

where ϵ is a small number.

Illustration of a barrier functions



Pink is the domain. The ellipsoids represent the shape of the hessian $\frac{\partial^2 J}{\partial \mathbf{x}^2}$ at different points on the domain. Green dot is $\mathbf{x}_a$.

Lecture slides are available via Github, links are on Moodle:

github.com/SergeiSa/Convex-Optimization