1 Projections by KKT

Let $\Omega = \{x \in \mathbb{R}^n : f_i(x) \le 0,\ h_j(x) = 0 \text{ for all } i, j\}$, with $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, p\}$. The projection of $u \in \mathbb{R}^n$ onto Ω is

$$\text{proj}_\Omega(u) = \arg\min_{x \in \Omega} \frac{1}{2} \|x - u\|^2 .$$

Each item below gives a set Ω and a formula for $\text{proj}_\Omega(u)$. For each item:

1. Write the constraints, and check convexity and Slater's condition.
2. Write the KKT conditions.
3. Find multipliers with which the given formula satisfies the KKT conditions.

Rigorously justify each conclusion.

Worked example 1.1

The nonnegative orthant $\Omega = \{x \in \mathbb{R}^n : x_i \ge 0 \text{ for all } i\}$, with $[\text{proj}_\Omega(u)]_i = \max\{u_i, 0\}$.

Solution.

Step 1: Slater's condition. Write $f_i(x) = -x_i$ for $i = 1, \dots, n$. Each f_i is affine, hence convex. At $x = (1, \dots, 1)$, every $f_i = -1 < 0$. So Slater's condition holds.

Step 2: KKT conditions. The gradients are $\nabla f_i = -e_i$. The KKT conditions, coordinate by coordinate, are

$$\begin{aligned} \text{stationarity:} \quad & x_i - u_i - y_i = 0, \\ \text{sign:} \quad & y_i \ge 0, \\ \text{complementary slackness:} \quad & y_i \cdot (-x_i) = 0, \\ \text{feasibility:} \quad & -x_i \le 0. \end{aligned}$$

Step 3: verify. Take $x_i = \max\{u_i, 0\}$ and $y_i = \max\{-u_i, 0\}$. Each coordinate gives two cases.

Case 1: $u_i \ge 0$. Then $x_i = u_i$ and $y_i = 0$. Stationarity reads $u_i - u_i - 0 = 0$. Since $y_i = 0$ and $x_i = u_i \ge 0$, the other three conditions hold.

Case 2: $u_i < 0$. Then $x_i = 0$ and $y_i = -u_i > 0$. Stationarity reads $0 - u_i + u_i = 0$. Since $x_i = 0$, complementary slackness and feasibility hold.

The problem is convex and Slater's condition holds, so $\max\{u, 0\}$ is the projection.

1. The interval $\Omega = \{x \in \mathbb{R} : a \le x \le b\}$ with $a < b$, and
$$\text{proj}_\Omega(u) = \min\{\max\{u, a\}, b\} = \begin{cases} a, & u < a, \\ u, & a \le u \le b, \\ b, & u > b. \end{cases}$$
2. The half-space $\Omega = \{x \in \mathbb{R}^n : a^T x \le b\}$, where $a \in \mathbb{R}^n$ is nonzero and $b \in \mathbb{R}$, and
$$\text{proj}_\Omega(u) = u - \frac{\max\{a^T u - b, 0\}}{\|a\|^2} a.$$
3. The ball $\Omega = \{x \in \mathbb{R}^n : \|x\| \le r\}$ with $r > 0$, and
$$\text{proj}_\Omega(u) = \begin{cases} u, & \|u\| \le r, \\ \dfrac{r}{\|u\|} u, & \|u\| > r. \end{cases}$$
4. **(G)** The affine set $\Omega = \{x \in \mathbb{R}^n : Ax = b\}$, where $A \in \mathbb{R}^{p \times n}$ has full row rank, and
$$\text{proj}_\Omega(u) = u - A^T (AA^T)^{-1} (Au - b).$$
The matrix AA^T is invertible.

2 Fixed-step projected gradient descent

Let $f_0 : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable with $\|\nabla f_0(x) - \nabla f_0(y)\| \le L\|x - y\|$ for all x, y, so that

$$f_0(y) \le f_0(x) + \nabla f_0(x)^T(y - x) + \frac{L}{2}\|y - x\|^2 \qquad \text{for all } x, y \in \mathbb{R}^n. \tag{Q}$$

Let $\Omega = \{x \in \mathbb{R}^n : f_i(x) \le 0,\ h_j(x) = 0 \text{ for all } i, j\}$, with $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, p\}$, satisfy Slater's condition, where every f_i is convex and continuously differentiable and every h_j is affine, and let $p^\star = \inf_{x\in\Omega} f_0(x) > -\infty$. Projected gradient descent with step size $\alpha > 0$ starts from $x_0 \in \Omega$ and sets

$$x_{k+1} = \mathrm{proj}_\Omega\big(x_k - \alpha\nabla f_0(x_k)\big).$$

Write $x_{k+1} = x_k - \alpha g_k$. Without constraints, $g_k = \nabla f_0(x_k)$. The projection problem is convex and satisfies Slater's condition, so its KKT conditions hold at x_{k+1}. Divided by α, they give multipliers $y \ge 0$ and z with $y_i f_i(x_{k+1}) = 0$ for all i and

$$\nabla f_0(x_k) + \sum_{i=1}^{m} y_i \nabla f_i(x_{k+1}) + \sum_{j=1}^{p} z_j \nabla h_j(x_{k+1}) = g_k. \tag{S}$$

The iterates satisfy the projection inequality and the descent lemma,

$$\big(\nabla f_0(x_k) - g_k\big)^T(w - x_{k+1}) \ge 0 \qquad \text{for all } w \in \Omega, \tag{P}$$

$$f_0(x_{k+1}) \le f_0(x_k) - \alpha\left(1 - \frac{L\alpha}{2}\right)\|g_k\|^2. \tag{DL}$$

1. **Projected step.** Starting from (S), the tangent inequality

   $$f_i(w) \ge f_i(x_{k+1}) + \nabla f_i(x_{k+1})^T(w - x_{k+1})$$

   of each convex f_i, and the identity $\nabla h_j(x_{k+1})^T(w - x_{k+1}) = h_j(w) - h_j(x_{k+1})$ of each affine h_j, prove (P).

2. **Descent lemma.** Prove (DL) in two steps. First, take $w = x_k$ in (P), which is allowed because every iterate lies in Ω, and use $x_k - x_{k+1} = \alpha g_k$ to show that

   $$\nabla f_0(x_k)^T(x_k - x_{k+1}) \ge \alpha\|g_k\|^2.$$

   Second, take $x = x_k$ and $y = x_{k+1}$ in (Q), and use this inequality.

3. **Stationarity.** Let $\mathscr{L}(x, y, z) = f_0(x) + \sum_i y_i f_i(x) + \sum_j z_j h_j(x)$ be the Lagrangian, and let y and z be the multipliers in (S). Prove that $\|\nabla_x \mathscr{L}(x_{k+1}, y, z)\| \le (1 + L\alpha)\|g_k\|$. Then let $\alpha = \frac{1}{L}$, and sum (DL) to conclude that some $x_{\text{best}} \in \{x_1, \dots, x_N\}$ has multipliers $y \ge 0$ and z with $y_i f_i(x_{\text{best}}) = 0$ for all i and

   $$\|\nabla_x \mathscr{L}(x_{\text{best}}, y, z)\| \le 2\sqrt{\frac{2L(f_0(x_0) - p^\star)}{N}}.$$

   Hint: by (S), $\nabla_x \mathscr{L}(x_{k+1}, y, z) = \nabla f_0(x_{k+1}) - \nabla f_0(x_k) + g_k$.

3 The barrier method

Let $a_1, \dots, a_m$ span $\mathbb{R}^n$, let $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$, and write $Ax < b$ when $a_i^T x < b_i$ for all i. For $\mu > 0$, the barrier function is

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

on $\{x : Ax < b\}$. Its Hessian does not depend on μ, and it is positive definite because the a_i span $\mathbb{R}^n$. The Newton step and the Newton decrement of F_μ at x are

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

Since F_μ is self-concordant, a full Newton step from a point with $\delta_\mu(x) < 1$ satisfies $A(x+v) < b$ and

$$\delta_\mu(x+v) \le \left(\frac{\delta_\mu(x)}{1 - \delta_\mu(x)}\right)^2. \tag{QC}$$

For every x with $Ax < b$ and every $0 < \mu' \le \mu$, decreasing μ to μ' gives

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

This follows from three bounds, which hold for every such x and μ' and every $u \in \mathbb{R}^n$,

$$\nabla F_\mu(x)^T u \le \delta_\mu(x) \sqrt{u^T \nabla^2 F_\mu(x) u}, \tag{CS}$$

$$\left| \sum_{i=1}^{m} \frac{a_i^T u}{b_i - a_i^T x} \right| \le \sqrt{m}\, \sqrt{u^T \nabla^2 F_\mu(x) u}, \tag{BT}$$

$$\nabla F_{\mu'}(x)^T u \le \left(\frac{\mu}{\mu'}\, \delta_\mu(x) + \left(\frac{\mu}{\mu'} - 1\right) \sqrt{m}\right) \sqrt{u^T \nabla^2 F_\mu(x) u}. \tag{DU}$$

The short-step method starts from x_0 with $Ax_0 < b$ and $\delta_{\mu_0}(x_0) \le \frac{1}{3}$, and sets $x_{k+1} = x_k + v_k$ and $\mu_{k+1} = \mu_k / (1 + \frac{1}{15\sqrt{m}})$, where v_k is the Newton step of F_{μ_k} at x_k.

1. **Decreasing μ at the center.** Let $a_i = -e_i$ and $b_i = 0$ for $i = 1, \dots, n$, so that $m = n$ and $F_\mu(x) = \frac{1}{\mu} c^T x - \sum_i \log x_i$ on $x > 0$, and let every $c_i > 0$. Find the minimizer of F_μ. At this point, compute $\delta_{\mu'}(x)$ for $0 < \mu' \le \mu$, and show that (DM) holds with equality.
2. **Short steps.** Starting from (QC) and (DM), prove that the short-step method keeps $Ax_k < b$ and $\delta_{\mu_k}(x_k) \le \frac{1}{3}$ for every k.
3. **Step count.** Let $0 < \epsilon < 1$. Using $\log(1+t) \ge \frac{t}{1+t}$ for $t \ge 0$, where log is the natural logarithm, prove that the short-step method has $\mu_k \le \epsilon \mu_0$ for every $k \ge (15\sqrt{m} + 1) \log \frac{1}{\epsilon}$.
4. **(G) The barrier term.** Let $Ax < b$. Prove (BT) by the Cauchy–Schwarz inequality in $\mathbb{R}^m$.
5. **(G) Decreasing μ along a direction.** Let $Ax < b$ and $0 < \mu' \le \mu$. Starting from (CS) and (BT), prove (DU).
6. **(G) Decreasing μ.** Let $Ax < b$, let $0 < \mu' \le \mu$, and let $u = \nabla^2 F_\mu(x)^{-1} \nabla F_{\mu'}(x)$. Prove (DM) by applying (DU) to this u.

4 Newton's method with linear equality constraints

Let $A \in \mathbb{R}^{p\times n}$ have full row rank, let $b \in \mathbb{R}^p$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable and strongly convex, so that $\nabla^2 f(x) \succ 0$ for every x. The problem is to minimize $f(x)$ subject to $Ax = b$. Since f is convex and the constraints are affine, x is a minimizer if and only if $r(x, z) = 0$ for some $z \in \mathbb{R}^p$, where

$$r(x,z) = \begin{bmatrix} \nabla f(x) + A^T z \\ Ax - b \end{bmatrix}.$$

At any $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^p$, the Newton step (v, w) for the equations $r(x, z) = 0$ solves

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = -r(x,z) = \begin{bmatrix} -\nabla f(x) - A^T z \\ b - Ax \end{bmatrix}, \tag{N}$$

and a full step moves from (x, z) to $(x+v, z+w)$. This system has exactly one solution. A feasible start has $Ax = b$, so the second block of the right side is zero. An infeasible start has $Ax \neq b$.

1. **Newton step.** Let $Ax = b$ and $z \in \mathbb{R}^p$. The Newton step minimizes the quadratic model over the steps that keep $Ax = b$,

   $$\text{minimize } f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v \text{ subject to } Av = 0.$$

   Show that the KKT conditions of this problem, with multiplier $z + w$ for $Av = 0$, are (N). Prove that $A(x + \alpha v) = b$ for every α.

2. **Uniqueness.** Let $x \in \mathbb{R}^n$. Prove that if $\nabla^2 f(x) v + A^T w = 0$ and $Av = 0$, then $v = 0$ and $w = 0$. Conclude that (N) has exactly one solution.

3. **Decrement.** Let $Ax = b$ and $z \in \mathbb{R}^p$, let (v, w) solve (N), and let $\delta \geq 0$ satisfy $\delta^2 = v^T \nabla^2 f(x) v$. Prove that $\delta^2 = -\nabla f(x)^T v$.

4. **(G) Infeasible start.** Starting from any $x_0 \in \mathbb{R}^n$ and $z_0 \in \mathbb{R}^p$, let

   $$x_{k+1} = x_k + \alpha_k v_k, \qquad z_{k+1} = z_k + \alpha_k w_k,$$

   where (v_k, w_k) solves (N) at (x_k, z_k) and $0 < \alpha_k \leq 1$. Prove that

   $$Ax_k - b = (1-\alpha_0)(1-\alpha_1)\cdots(1-\alpha_{k-1})(Ax_0 - b) \qquad \text{for every } k \geq 1.$$

   Conclude that once some $\alpha_j = 1$, every later iterate satisfies $Ax_k = b$.

5. **(G) Residual along the step.** Let $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^p$ with $r(x, z) \neq 0$. Let (v, w) solve (N), and let

   $$\varphi(\alpha) = r(x + \alpha v,\, z + \alpha w) \qquad \text{for } \alpha \in \mathbb{R}.$$

   Prove that $\varphi'(0) = -r(x, z)$. Conclude that $\|\varphi(\alpha)\|$ is differentiable at $\alpha = 0$, with

   $$\frac{d}{d\alpha} \|\varphi(\alpha)\| \Big|_{\alpha=0} = -\|r(x,z)\|.$$

6. **(G) Residual test.** Let $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^p$ with $r(x, z) \neq 0$, and let $v \in \mathbb{R}^n$ and $w \in \mathbb{R}^p$ satisfy

   $$\frac{d}{d\alpha} \|r(x + \alpha v,\, z + \alpha w)\| \Big|_{\alpha=0} = -\|r(x,z)\|.$$

   Let $0 < c_1 < 1$. Prove that there is $\bar{\alpha} > 0$ such that

   $$\|r(x + \alpha v,\, z + \alpha w)\| \leq (1 - c_1\alpha)\, \|r(x,z)\| \qquad \text{for all } 0 < \alpha \leq \bar{\alpha}.$$

   Hint: let $\psi(\alpha) = \|r(x + \alpha v, z + \alpha w)\|$, so that $\psi'(0) = -\psi(0) < 0$. By the definition of the derivative, $\frac{\psi(\alpha) - \psi(0)}{\alpha} \to -\psi(0)$ as $\alpha \to 0^+$. Choose $\bar{\alpha}$ so that this quotient is at most $-c_1\psi(0)$ for $0 < \alpha \leq \bar{\alpha}$.