Duality

- Optimality conditions - Complementary slackness

Suppose that the primal and dual optimal values are attained and equal (so, in particular, strong duality holds). Let $\mathbf{x}^*$ be a primal optimal and $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ be a dual optimal point. This means that

$$\begin{aligned}
f_0(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}) \right) \\
&\leq f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}^*) \leq f_0(\mathbf{x}^*).
\end{aligned}$$

The first line states that the optimal duality gap is zero, and the second line is the definition of the dual function. The third line follows since the infimum of the Lagrangian over $\mathbf{x}$ is less than or equal to its value at $\mathbf{x} = \mathbf{x}^*$. The last inequality follows from $\lambda_i^* \geq 0$, $f_i(\mathbf{x}^*) \leq 0$, $i = 1, \cdots, m$, and $h_i(\mathbf{x}^*) = 0$, $i = 1, \cdots, p$. We conclude that the two inequalities in this chain hold with equality.

Duality

- Optimality conditions - Complementary slackness

We can draw several interesting conclusions from this. For example, since the inequality in the third line is an equality, we conclude that $\mathbf{x}^*$ minimizes $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ over $\mathbf{x}$. (The Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu})^*$ can have other minimizers; $\mathbf{x}^*$ is simply a minimizer.)

Another important conclusion is that

$$\sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) = 0.$$

Since each term in this sum is nonpositive, we conclude that

$$\lambda_i^* f_i(\mathbf{x}^*) = 0, \quad i = 1, \cdots, m. \tag{17}$$

Duality

- Optimality conditions - Complementary slackness

This condition is known as *complementary slackness*; it holds for any primal optimal $\mathbf{x}^*$ and any dual optimal $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ (when strong duality holds). We can express the complementary slackness condition as

$$\lambda_i^* > 0 \Longrightarrow f_i(\mathbf{x}^*) = 0,$$

or, equivalently,

$$f_i(\mathbf{x}^*) < 0 \Longrightarrow \lambda_i^* = 0.$$

Roughly speaking, this means the ith optimal Lagrange multiplier is zero unless the ith constraint is active at the optimum.

Duality

- Optimality conditions - KKT optimality conditions

We now assume that the functions $f_0, \cdots, f_m$, $h_1, \cdots, h_p$ are differentiable (and therefore have open domains), but we make no assumptions yet about convexity.

KKT conditions for nonconvex problems

Let $\mathbf{x}^*$ and $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ be any primal and dual optimal points with zero duality gap. Since $\mathbf{x}^*$ minimizes $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ over $\mathbf{x}$, it follows that its gradient must vanish at $\mathbf{x}^*$, *i.e.*,

$$\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{n} \nu_i^* \nabla h_i(\mathbf{x}^*) = \mathbf{0}.$$

Duality

- Optimality conditions - KKT optimality conditions

Thus we have

$$
\begin{aligned}
f_i(\mathbf{x}^*) &\leq 0, \quad i = 1, \cdots, m \\
h_i(\mathbf{x}^*) &= 0, \quad i = 1, \cdots, p \\
\lambda_i^* &\geq 0, \quad i = 1, \cdots, m \\
\lambda_i^* f_i(\mathbf{x}^*) &= 0, \quad i = 1, \cdots, m
\end{aligned} \tag{18}
$$

$$
\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(\mathbf{x}^*) = \mathbf{0},
$$

which are called the *Karush-Kuhn-Tucker* (KKT) conditions.

To summarize, for *any* optimization problem with differentiable objective and constraint functions for which strong duality obtains, any pair of primal and dual optimal points must satisfy the KKT conditions (18).

Duality

- Optimality conditions - KKT optimality conditions

KKT conditions for convex problems

When the primal problem is convex, the KKT conditions are also sufficient for the points to be primal and dual optimal. In other words, if f_i are convex and h_i are affine, and $\tilde{\mathbf{x}}$, $\tilde{\boldsymbol{\lambda}}$, $\tilde{\boldsymbol{\nu}}$ are any points that satisfy the KKT conditions

$$
\begin{aligned}
f_i(\tilde{\mathbf{x}}) &\leq 0, \quad i = 1, \cdots, m \\
h_i(\tilde{\mathbf{x}}) &= 0, \quad i = 1, \cdots, p \\
\tilde{\lambda}_i &\geq 0, \quad i = 1, \cdots, m \\
\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) &= 0, \quad i = 1, \cdots, m \\
\nabla f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i \nabla h_i(\tilde{\mathbf{x}}) &= \mathbf{0},
\end{aligned}
$$

then $\tilde{\mathbf{x}}$ and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ are primal and dual optimal, with zero duality gap.

Duality

- Optimality conditions - KKT optimality conditions

Proof. Note that the first two conditions state that $\tilde{\mathbf{x}}$ is primal feasible. Since $\tilde{\lambda}_i \geq 0$, $L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ is convex in $\mathbf{x}$; the last KKT condition states that its gradient with respect to $\mathbf{x}$ vanishes at $\mathbf{x} = \tilde{\mathbf{x}}$, so it follows that $\tilde{\mathbf{x}}$ minimizes $L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ over $\mathbf{x}$. From this we conclude that

$$\begin{aligned} g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) &= L(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) \\ &= f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{\mathbf{x}}) \\ &= f_0(\tilde{\mathbf{x}}), \end{aligned}$$

where in the last line we use $h_i(\tilde{\mathbf{x}}) = 0$ and $\tilde{\boldsymbol{\lambda}}_i f_i(\tilde{\mathbf{x}}) = 0$. This shows that $\tilde{\mathbf{x}}$ and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ have zero duality gap, and therefore are primal and dual optimal. In summary, for any *convex* optimization problem with differentiable objective and constraint functions, any points that satisfy the KKT conditions are primal and dual optimal, and have zero duality gap.

Duality

- Optimality conditions - KKT optimality conditions

$$
\begin{tikzcd}
\text{SCQ} \arrow[r, Rightarrow] & \text{Strong duality} \arrow[r, Rightarrow, shift right] \arrow[d, Leftrightarrow] & \text{KKT conditions} \arrow[l, Rightarrow, shift right, "\text{CP}"'] \\
 & \text{Minimum/Maximum} &
\end{tikzcd}
$$

Duality

- Optimality conditions - Solving the primal problem via the dual

If strong duality holds and a dual optimal solution $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ exists, then any primal optimal point is also a minimizer of $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$. This fact sometimes allows us to compute a primal optimal solution from a dual optimal solution.

More precisely, suppose we have strong duality and an optimal $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ is known. Suppose that the minimizer of $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, *i.e.*, the solution of

$$\text{minimize} \quad f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}), \tag{19}$$

is unique. (For a convex problem this occurs, for example, if $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ is a strictly convex function of $\mathbf{x}$.) Then if the solution of (19) is primal feasible, it must be primal optimal; if it is not primal feasible, then no primal optimal point can exist, *i.e.*, we can conclude that the primal optimum is not attained. This observation is interesting when the dual problem is easier to solve than the primal problem, for example, because it can be solved analytically, or has some special structure that can be exploited.

Examples.

Duality

- Solving the primal problem via the dual – Three types of reformulation

Simple equivalent reformulations of a problem can lead to very different dual problems. We consider the following types of reformulations:

- Introducing new variables and associated equality constraints.
- Replacing the objective with an increasing function of the original objective.
- Making explicit constraints implicit, i.e., incorporating them into the domain of the objective.

Examples.

Duality

- Theorems of alternatives - Weak alternatives

In this section we apply Lagrange duality theory to the problem of determining feasibility of a system of inequalities and equalities

$$f_i(\mathbf{x}) \leq 0, \quad i = 1, \cdots, m, \quad h_i(\mathbf{x}) = 0, \quad i = 1, \cdots, p. \tag{20}$$

We can think of (20) as the standard problem (1), with objective $f_0 = 0$, *i.e.*,

$$\begin{aligned} \min_{\mathbf{x}} \quad & 0 \\ s.t. \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \cdots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \cdots, p. \end{aligned} \tag{21}$$

This problem has optimal value

$$p^* = \begin{cases} 0 & (20) \text{ is feasible} \\ \infty & (20) \text{ is infeasible,} \end{cases} \tag{22}$$

so solving the optimization problem (21) is the same as solving the inequality system (20).

Duality

- Theorems of alternatives - Weak alternatives

The dual function

We associate with the inequality system (20) the dual function

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} \left(\sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \right),$$

which is the same as the dual function for the optimization problem (21). Since $f_0 = 0$, the dual function is positive homogeneous in $(\boldsymbol{\lambda}, \boldsymbol{\nu})$: For α, $g(\alpha\boldsymbol{\lambda}, \alpha\boldsymbol{\nu}) = \alpha g(\boldsymbol{\lambda}, \boldsymbol{\nu})$. The dual problem associated with (21) is to maximize $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ subject to $\boldsymbol{\lambda} \geq \mathbf{0}$. Since g is homogeneous, the optimal value of this dual problem is given by

$$d^* = \begin{cases} \infty & \boldsymbol{\lambda} \geq \mathbf{0}, \quad g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > 0 \text{ is feasible} \\ 0 & \boldsymbol{\lambda} \geq \mathbf{0}, \quad g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > 0 \text{ is infeasible.} \end{cases} \tag{23}$$

Duality

- Theorems of alternatives - Weak alternatives

Weak duality tells us that $d^* \leq p^*$. Combining this fact with (22) and (23) yields the following: If the inequality system

$$\boldsymbol{\lambda} \geq \mathbf{0}, \quad g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > 0 \tag{24}$$

is feasible (which means $d^* = \infty$), then the inequality system (20) is infeasible (since we then have $p^* = \infty$). Indeed, we can interpret any solution $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ of the inequalities (24) as a *proof* or *certificate* of infeasibility of the system (20).

We can restate this implication in terms of feasibility of the original system: If the original inequality system (20) is feasible, then the inequality system (24) must be infeasible. We can interpret an $\mathbf{x}$ which satisfies (20) as a certificate establishing infeasibility of the inequality system (24).

Two systems of inequalities (and equalities) are called *weak alternatives* if at most one of the two is feasible. Thus, the systems (20) and (24) are weak alternatives. This is true whether or not the inequalities (20) are convex (i.e., f_i convex, h_i affine); moreover, the alternative inequality system (24) is always convex (*i.e.*, g is concave and the constraints $\lambda_i \geq 0$ are convex).

Duality

- Theorems of alternatives - Weak alternatives

Strict inequalities

We can also study feasibility of the *strict* inequality system

$$f_i(\mathbf{x}) < 0, \quad i = 1, \cdots, m, \quad h_i(\mathbf{x}) = 0, \quad i = 1, \cdots, p. \tag{25}$$

With g defined as for the nonstrict inequality system, we have the alternative inequality system

$$\boldsymbol{\lambda} \geq \mathbf{0}, \quad \boldsymbol{\lambda} \neq \mathbf{0}, \quad g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \geq 0. \tag{26}$$

We can show directly that (25) and (26) are weak alternatives. Suppose there exists an $\tilde{\mathbf{x}}$ with $f_i(\tilde{\mathbf{x}}) < 0$, $h_i(\tilde{\mathbf{x}}) = 0$. Then for any $\boldsymbol{\lambda} \geq \mathbf{0}$, $\boldsymbol{\lambda} \neq \mathbf{0}$, and $\boldsymbol{\nu}$,

$$\lambda_1 f_1(\tilde{\mathbf{x}}) + \cdots + \lambda_m f_m(\tilde{\mathbf{x}}) + \nu_1 h_1(\tilde{\mathbf{x}}) + \cdots + \nu_p h_p(\tilde{\mathbf{x}}) < 0.$$

Duality

- Theorems of alternatives - Weak alternatives

It follows that

$$
\begin{aligned}
g(\boldsymbol{\lambda}, \boldsymbol{\nu}) &= \inf_{\mathbf{x} \in \mathcal{D}} \left(\sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \right) \\
&\leq \sum_{i=1}^{m} \lambda_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{\mathbf{x}}) \\
&< 0.
\end{aligned}
$$

Therefore, feasibility of (25) implies that there does not exist $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ satisfying (26).

Thus, we can prove infeasibility of (25) by producing a solution of the system (26); we can prove infeasibility of (26) by producing a solution of the system (25).

Duality

- Theorems of alternatives - Strong alternatives

When the original inequality system is convex, *i.e.*, f_i are convex and h_i are affine, and some type of constraint qualification holds, then the pairs of weak alternatives described above are *strong alternatives*, which means that *exactly one* of the two alternatives holds. In other words, each of the inequality systems is feasible if and only if the other is infeasible.

In this section we assume that f_i are convex and h_i are affine, so the inequality system (20) can be expressed as

$$f_i(\mathbf{x}) \leq 0, \quad i = 1, \cdots, m, \quad \mathbf{A}\mathbf{x} = \mathbf{b},$$

where $\mathbf{A} \in \mathbb{R}^{p \times n}$.

Duality

- Theorems of alternatives - Strong alternatives

Strict inequalities

We first study the strict inequality system

$$f_i(\mathbf{x}) < 0, \quad i = 1, \cdots, m, \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \tag{27}$$

and its alternative

$$\boldsymbol{\lambda} \geq \mathbf{0}, \quad \boldsymbol{\lambda} \neq \mathbf{0}, \quad g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \geq 0. \tag{28}$$

We need one technical condition: There exists an $\mathbf{x} \in \mathbf{relint}\ \mathbf{D}$ with $\mathbf{A}\mathbf{x} = \mathbf{b}$. In other words we not only assume that the linear equality constraints are consistent, but also that they have a solution in **relint** $\mathcal{D}$. (Very often $\mathcal{D} = \mathbb{R}^n$, so the condition is satisfied if the equality constraints are consistent.) Under this condition, exactly one of the inequality systems (27) and (28) is feasible. In other words, the inequality systems (27) and (28) are strong alternatives.

Duality

- Theorems of alternatives - Strong alternatives

We will establish this result by considering the related optimization problem

$$
\begin{aligned}
\min_{\mathbf{x},s} \quad & s \\
s.t. \quad & f_i(\mathbf{x}) - s \leq 0, \quad i = 1, \cdots, m \\
& Ax = b
\end{aligned}
\tag{29}
$$

with domain $\mathcal{D} \times \mathbb{R}$. The optimal value p^* of this problem is negative if and only if there exists a solution to the strict inequality system (27).

The Lagrange dual function for the problem (29) is

$$
\inf_{\mathbf{x}\in\mathcal{D},s} \left(s + \sum_{i=1}^{m} \lambda_i (f_i(\mathbf{x}) - s) + \boldsymbol{\nu}^T (\mathbf{A}\mathbf{x} - \mathbf{b}) \right) = \begin{cases} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) & \mathbf{1}^T \boldsymbol{\lambda} = 1 \\ -\infty & \text{otherwise.} \end{cases}
$$

Therefore we can express the dual problem of (29) as

$$
\begin{aligned}
\max_{\boldsymbol{\lambda},\boldsymbol{\nu}} \quad & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\
s.t. \quad & \boldsymbol{\lambda} \geq \mathbf{0}, \quad \mathbf{1}^T \lambda = 1.
\end{aligned}
$$

Duality

- Theorems of alternatives - Strong alternatives

Now we observe that Slater's condition holds for the problem (29). By the hypothesis there exists an $\tilde{\mathbf{x}} \in \textbf{relint}\ \mathcal{D}$ with $\mathbf{A}\tilde{\mathbf{x}} = \mathbf{b}$. Choosing any $\tilde{s} > \max_i f_i(\tilde{\mathbf{x}})$ yields a point $(\tilde{\mathbf{x}}, \tilde{s})$ which is strictly feasible for (29). Therefore we have $d^* = p^*$, and the dual optimum d^* is attained. In other words, there exist $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ such that

$$g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = p^* \quad \boldsymbol{\lambda}^* \geq \mathbf{0}, \quad \mathbf{1}^T \boldsymbol{\lambda}^* = 1. \tag{30}$$

Now suppose that the strict inequality system (27) is infeasible, which means that $p^* \geq 0$. Then $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ from (29) satisfy the alternate inequality system (28). Similarly, if the alternate inequality system (28) is feasible, then $d^* = p^* \geq 0$, which shows that the strict inequality system (27) is infeasible. Thus, the inequality systems (27) and (28) are strong alternatives; each is feasible iff the other is not.

Duality

- Theorems of alternatives - Strong alternatives

Nonstrict inequalities

We now consider the nonstrict inequality system

$$f_i(\mathbf{x}) \leq 0, \quad i = 1, \cdots, m, \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \tag{31}$$

and its alternative

$$\boldsymbol{\lambda} \geq \mathbf{0}, \quad g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > 0 \tag{32}$$

We will show these are strong alternatives, if the following conditions hold: There exists an $\mathbf{x} \in \mathbf{relint}\ \mathcal{D}$ with $\mathbf{A}\mathbf{x} = \mathbf{b}$, and the optimal value p^* of (29) is attained. This holds, for example, if $\mathcal{D} = \mathbb{R}^n$ and $\max_i f_i(\mathbf{x}) \to \infty$ as $\mathbf{x} \to \infty$. With these assumptions, as in the strict case, that $p^* = d^*$, and that both the primal and dual optimal values are attained. Now suppose that the nonstrict inequality system (31) is infeasible, which means that $p^* > 0$. (Here we use the assumption that the primal optimal value is attained.) Then $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ from (30) satisfy the alternate inequality system (32). Thus, the inequality systems (31) and (32) are strong alternatives; each is feasible iff the other is not.

Duality

- Theorems of alternatives - Strong alternatives

Examples: Linear inequalities, Intersection of ellipsoids, Farkas' lemma, Arbitrage-free bounds on price

Chapter 7. Constrained Optimization

- Algorithms for constrained optimization
- Frank-Wolfe method
- Alternating direction method (ADM)
- Linearized alternating direction method (LADM)
- Proximal Linearized alternating direction method (PLADM)
- Coordinate descent and block coordinate descent

Algorithms for constrained optimization

- Projections

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ s.t. \quad & \mathbf{x} \in \Omega. \end{aligned}$$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)},$$

where $\mathbf{d}^{(k)}$ is typically a function of $\nabla f(\mathbf{x}^{(k)})$, may not work. A simple modification involves the introduction of a projection:

$$\mathbf{x}^{(k+1)} = \mathcal{P}_\Omega(\mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}).$$

The projected version of the gradient algorithm has the form

$$\mathbf{x}^{(k+1)} = \mathcal{P}_\Omega(\mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})).$$

We refer to the above as the *projected gradient algorithm*.

Algorithms for constrained optimization

- Projections - Example

$$
\begin{aligned}
\min \quad & \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} \\
s.t. \quad & \|\mathbf{x}\|^2 = 1,
\end{aligned}
$$

where $\mathbf{Q} = \mathbf{Q}^T \succ \mathbf{0}$. Suppose that we apply a fixed-step-size projected gradient algorithm to this problem.

Algorithms for constrained optimization

- Projected Gradient Methods with Linear Constraints

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ s.t. \quad & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, $m < n$, $\operatorname{rank} \mathbf{A} = m$, $\mathbf{b} \in \mathbb{R}^m$.

$$\mathbf{P} = \mathbf{I}_n - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A},$$

Lemma 1. *Let $\mathbf{v} \in \mathbb{R}^n$. Then, $\mathbf{P}\mathbf{v} = \mathbf{0}$ if and only if $\mathbf{v} \in \mathcal{R}(\mathbf{A}^T)$. In other words, $\mathcal{N}(\mathbf{P}) = \mathcal{R}(\mathbf{A}^T)$. Moreover, $\mathbf{A}\mathbf{v} = \mathbf{0}$ if and only if $\mathbf{v} \in \mathcal{R}(\mathbf{P})$; that is, $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{P})$.*

Proposition 2. *Let $\mathbf{x}^* \in \mathbb{R}^n$ be a feasible point. Then, $\mathbf{P}\nabla f(\mathbf{x}^*) = \mathbf{0}$ if and only if $\mathbf{x}^*$ satisfies the Lagrange condition.*

$$\mathbf{x}^{(k)} \in \Omega \Longrightarrow \mathbf{x}^{(k+1)} = \mathcal{P}_\Omega(\mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})) = \mathbf{x}^{(k)} - \alpha_k \mathbf{P}\nabla f(\mathbf{x}^{(k)}).$$

Algorithms for constrained optimization

- Projected Gradient Methods with Linear Constraints

Proposition 1. *In projected gradient algorithm, if* $\mathbf{x}^{(0)}$ *is feasible, then each* $\mathbf{x}^{(k)}$ *is feasible; that is, for each* $k > 0, \mathbf{A}\mathbf{x}^{(k)} = \mathbf{b}$.

Proposition 2. $-\mathbf{P}\nabla f(\mathbf{x}^{(k)})$ *points in the direction of maximum rate of decrease of* f *at* $\mathbf{x}^{(k)}$ *along the surface defined by* $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Proposition 3. *If* $-\mathbf{P}\nabla f(\mathbf{x}^{(k)}) \neq \mathbf{0}$ *then it is a descent direction.*

Algorithms for constrained optimization

- Equality constrained convex quadratic minimization

$$\min_{\mathbf{x}} \ \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{q}^T\mathbf{x} + r,$$
$$s.t. \ \mathbf{A}\mathbf{x} = \mathbf{b},$$

where $\mathbf{P} \in \mathbb{S}_+^n$. The optimality conditions are:

$$\mathbf{A}\mathbf{x}^* = \mathbf{b}, \quad \mathbf{P}\mathbf{x}^* + \mathbf{q} + \mathbf{A}^T\boldsymbol{\nu}^* = \mathbf{0},$$

which can be written as

KKT matrix

$$\boxed{\boxed{\begin{pmatrix} \mathbf{P} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{pmatrix}} \begin{pmatrix} \mathbf{x}^* \\ \boldsymbol{\nu}^* \end{pmatrix} = \begin{pmatrix} -\mathbf{q} \\ \mathbf{b} \end{pmatrix}.}$$

KKT system

Algorithms for constrained optimization

- Equality constrained convex quadratic minimization

Nonsingularity of the KKT matrix

Recall our assumption that $\mathbf{P} \in \mathbb{S}_+^n$ and $\operatorname{rank} \mathbf{A} = p < n$. There are several conditions equivalent to nonsingularity of the KKT matrix:

- $\mathcal{N}(\mathbf{P}) \in \mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$, i.e., $\mathbf{P}$ and $\mathbf{A}$ have no nontrivial common nullspace.
- $\mathbf{Ax} = \mathbf{0}$, $\mathbf{x} \neq \mathbf{0} \Longrightarrow \mathbf{x}^T\mathbf{Px} > 0$, i.e., $\mathbf{P}$ is positive definite on the nullspace of $\mathbf{A}$.
- $\mathbf{F}^T\mathbf{PF} \succ \mathbf{0}$, where $\mathbf{F} \in \mathbb{R}^{n\times(n-p)}$ is a matrix for which $\mathcal{R}(\mathbf{F}) = \mathcal{N}(\mathbf{A})$.

As an important special case, we note that if $\mathbf{P} \succ \mathbf{0}$, the KKT matrix must be nonsingular.

Algorithms for constrained optimization

- Equality constrained convex quadratic minimization

Solving KKT systems

$$\begin{pmatrix} \mathbf{H} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \mathbf{w} \end{pmatrix} = - \begin{pmatrix} \mathbf{g} \\ \mathbf{h} \end{pmatrix}. \tag{2}$$

Here we assume $\mathbf{H} \in \mathbb{S}_+^n$ and $\mathbf{A} \in \mathbb{R}^{p \times n}$ with rank $\mathbf{A} = p < n$.

Solving KKT systems

One straightforward approach is to simply solve the KKT system (1), which is a set of $n + p$ linear equations in $n + p$ variables. The KKT matrix is symmetric, but not positive definite, so a good way to do this is to use an LDL^T factorization. If no structure of the matrix is exploited, the cost is $(1/3)(n+p)^3$ flops. This can be a reasonable approach when the problem is small (i.e., n and p are not too large), or when $\mathbf{A}$ and $\mathbf{H}$ are sparse.

Algorithms for constrained optimization

- Equality constrained convex quadratic minimization

Solving KKT system via elimination

We start by describing the simplest case, in which $\mathbf{H} \succ \mathbf{0}$. Starting from the first of the KKT equations

$$\mathbf{H}\mathbf{v} + \mathbf{A}^T\mathbf{b} = -\mathbf{g}, \quad \mathbf{A}\mathbf{v} = -\mathbf{h},$$

we solve for $\mathbf{v}$ to obtain

$$\mathbf{v} = -\mathbf{H}^{-1}(\mathbf{g} + \mathbf{A}^T\mathbf{w}).$$

Substituting this into the second KKT equation yields $\mathbf{A}\mathbf{H}^{-1}(\mathbf{g} + \mathbf{A}^T\mathbf{w}) = \mathbf{h}$, so we have

$$\mathbf{w} = (\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^T)^{-1}(\mathbf{h} - \mathbf{A}\mathbf{H}^{-1}\mathbf{g}).$$

These formulae give us a method for computing $\mathbf{v}$ and $\mathbf{w}$.

Algorithms for constrained optimization

- Equality constrained convex quadratic minimization

Solving KKT system by block elimination.
Given: KKT system with $\mathbf{H} \succ \mathbf{0}$.

1. Form $\mathbf{H}^{-1}\mathbf{A}^T$ and $\mathbf{H}^{-1}\mathbf{g}$.
2. Form Schur complement $\mathbf{S} = -\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^T$.
3. Determine $\mathbf{w}$ by solving $\mathbf{S}\mathbf{w} = \mathbf{A}\mathbf{H}^{-1}\mathbf{g} - \mathbf{h}$.
4. Determine $\mathbf{v}$ by solving $\mathbf{H}\mathbf{v} = -\mathbf{A}^T\mathbf{w} - \mathbf{g}$.

Algorithms for constrained optimization

- Equality constrained convex quadratic minimization

Step 1 can be done by a Cholesky factorization of $\mathbf{H}$, followed by $p+1$ solves, which costs $f + (p+1)s$, where f is the cost of factoring $\mathbf{H}$ and s is the cost of an associated solve. Step 2 requires a $p \times n$ by $n \times p$ matrix multiplication. If we exploit no structure in this calculation, the cost is $p^2 n$ flops. (Since the result is symmetric, we only need to compute the upper triangular part of S.) In some cases special structure in $\mathbf{A}$ and $\mathbf{H}$ can be exploited to carry out step 2 more efficiently. Step 3 can be carried out by Cholesky factorization of $-\mathbf{S}$, which costs $(1/3)p^3$ flops if no further structure of $\mathbf{S}$ is exploited. Step 4 can be carried out using the factorization of $\mathbf{H}$ already calculated in step 1, so the cost is $2np + s$ flops. The total flop count, assuming that no structure is exploited in forming or factoring the Schur complement, is

$$f + ps + p^2 n + (1/3)p^3$$

flops (keeping only dominant terms). If we exploit structure in forming or factoring $\mathbf{S}$, the last two terms are even smaller.

Algorithms for constrained optimization

- Equality constrained convex quadratic minimization

If $\mathbf{H}$ can be factored efficiently, then block elimination gives us a flop count advantage over directly solving the KKT system using an LDL^T factorization. For example, if $\mathbf{H}$ is diagonal (which corresponds to a separable objective function), we have $f = 0$ and $s = n$, so the total cost is $p^2n + (1/3)p^3$ flops, which grows only linearly with n. If $\mathbf{H}$ is banded with bandwidth $k \ll n$, then $f = nk^2$, $s = 4nk$, so the total cost is around $nk^2 + 4nkp + p^2n + (1/3)p^3$ which still grows only linearly with n. Other structures of $\mathbf{H}$ that can be exploited are block diagonal (which corresponds to block separable objective function), sparse, or diagonal plus low rank.

Examples: Equality constrained analytic center, Minimum length piecewise-linear curve subject to equality constraints, Locally linear embedding (LLE)

Algorithms for constrained optimization

- Equality constrained convex quadratic minimization

Elimination with singular H

The block elimination method described above obviously does not work when $\mathbf{H}$ is singular, but a simple variation on the method can be used in this more general case. The more general method is based on the following result: The KKT matrix is nonsingular if and only $\mathbf{H} + \mathbf{A}^T\mathbf{Q}\mathbf{A} \succ \mathbf{0}$ for some $\mathbf{Q} \succeq \mathbf{0}$, in which case, $\mathbf{H} + \mathbf{A}^T\mathbf{Q}\mathbf{A} \succ \mathbf{0}$ for all $\mathbf{Q} \succ \mathbf{0}$. We conclude, for example, that if the KKT matrix is nonsingular, then $\mathbf{H} + \mathbf{A}^T\mathbf{A} \succ \mathbf{0}$.

Let $\mathbf{Q} \succeq \mathbf{0}$ be a matrix for which $\mathbf{H} + \mathbf{A}^T\mathbf{Q}\mathbf{A} \succ \mathbf{0}$. Then the KKT system (2) is equivalent to

$$\begin{pmatrix} \mathbf{H} + \mathbf{A}^T\mathbf{Q}\mathbf{A} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \mathbf{w} \end{pmatrix} = -\begin{pmatrix} \mathbf{g} + \mathbf{A}^T\mathbf{Q}\mathbf{h} \\ \mathbf{h} \end{pmatrix}, \tag{3}$$

which can be solved using elimination since $\mathbf{H} + \mathbf{A}^T\mathbf{Q}\mathbf{A} \succ \mathbf{0}$.

Examples: Equality constrained analytic centering, Optimal network flow, Optimal network flow, Analytic center of a linear matrix inequality

Algorithms for constrained optimization

- Eliminating equality constraints

One general approach to solving the equality constrained problem is to elimınte the equality constraints and then solve the resulting unconstrained problem using methods for unconstrained minimization. We first find a matrix $\mathbf{F} \in \mathbb{R}^{n\times(n-p)}$ and vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ that parameterize the (affine) feasible set:

$$\{\mathbf{x}|\mathbf{A}\mathbf{x} = \mathbf{b}\} = \{\mathbf{F}\mathbf{z} + \hat{\mathbf{x}}|\mathbf{z} \in \mathbb{R}^{n-p}\}.$$

Here $\hat{bx}$ can be chosen as any particular solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{F} \in \mathbb{R}^{n\times(n-p)}$ is any matrix whose range is the nullspace of $\mathbf{A}$. We then form the reduced or eliminated optimization problem:

$$\min_{\mathbf{z}} \ \tilde{f}(\mathbf{z}) \triangleq f(\mathbf{F}\mathbf{z} + \hat{\mathbf{x}}), \tag{4}$$

which is an unconstrained problem with variable $\mathbf{z} \in \mathbb{R}^{n-p}$. From its solution $\mathbf{z}^*$, we can find the solution of the equality constrained problem as $\mathbf{x}^* = \mathbf{F}\mathbf{z}^* + \hat{\mathbf{x}}$.

Algorithms for constrained optimization

- Eliminating equality constraints

Example: Optimal allocation with resource constraint. We consider the problem:

$$
\begin{aligned}
\min_{\mathbf{x}} \ & \sum_{i=1}^{n} f_i(x_i), \\
s.t. \ & \sum_{i=1}^{n} x_i = b,
\end{aligned}
$$

where the functions $f_i : \mathbb{R} \to \mathbb{R}$ are convex and twice differentiable, and $b \in \mathbb{R}$ is a problem parameter. We interpret this as the problem of optimally allocating a single resource, with a fixed total amount b (the budget) to n otherwise independent activities.