Local First-Order Optimality Conditions

- Karush-Kuhn-Tucker (KKT) conditions

Theorem 2 (Karush-Kuhn-Tucker). *Suppose that* $\mathbf{x}_0$ *is a local minimizer of* (P)*, and the constraint qualification* $\mathcal{C}_l(\mathbf{x}_0)^* = \mathcal{C}_t(\mathbf{x}_0)^*$ *is fulfilled. Then there exist vectors* $\boldsymbol{\lambda} \in \mathbb{R}_+^m$ *and* $\boldsymbol{\mu} \in \mathbb{R}^p$ *such that*

$$\nabla f(\mathbf{x}_0) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}_0) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}_0) = \mathbf{0} \text{ and}$$

$$\lambda_i g_i(\mathbf{x}_0) = 0 \text{ for } i = 1, ..., m.$$

Proof. If $\mathbf{x}_0$ is a local minimizer of (P), it follows from Lemma 1 with the help of the presupposed constraint qualification that

$$\nabla f(\mathbf{x}_0) \in \mathcal{C}_t(\mathbf{x}_0)^* = \mathcal{C}_l(\mathbf{x}_0)^*;$$

Lemma 2 yields $\mathcal{C}_l(\mathbf{x}_0) \cap \mathcal{C}_{dd}(\mathbf{x}_0) = \emptyset$ and the latter together with Proposition 1 gives the result. $\square$

Local First-Order Optimality Conditions

- Karush-Kuhn-Tucker (KKT) conditions

For $\mathbf{x} = (x_1, x_2)^\top \in \mathbb{R}^2$ whether the feasible points $\mathbf{x}_0 := (-1, 0)^\top$ and $\tilde{\mathbf{x}}_0 := (0, 1)^\top$ are local minimizers of consider the problem

$$\begin{aligned} \min_{\mathbf{x}} \ & f(\mathbf{x}) := x_1 + x_2, \\ s.t. \ & -x_1^3 + x_2 \leq 1, \\ & x_1 \leq 1, -x_2 \leq 0. \end{aligned}$$

Local First-Order Optimality Conditions

- Constraint Qualifications

The condition $\mathcal{C}_l(\mathbf{x}_0)^* = \mathcal{C}_t(\mathbf{x}_0)^*$ is very abstract, extremely general, but not easily verifiable. Therefore, for practical problems, we will try to find regularity assumptions called *constraint qualifications* (CQ) which are more specific, easily verifiable, but also somewhat restrictive.

Assuming that we only have inequality constraints.
(GCQ) Guignard Constraint Qualification: $\mathcal{C}_l(\mathbf{x}_0)^* = \mathcal{C}_t(\mathbf{x}_0)^*$.
(ACQ) Abadie Constraint Qualification: $\mathcal{C}_l(\mathbf{x}_0) = \mathcal{C}_t(\mathbf{x}_0)$.
(SCQ) Slater Constraint Qualification: The functions g_i are convex for all $i \in \mathcal{I}$ and

$$\exists \tilde{\mathbf{x}} \in \mathcal{F},\ g_i(\tilde{\mathbf{x}}) < 0 \text{ for } i \in \boxed{\mathcal{I}_1}.$$

$\mathcal{I}_1$ is the index set of nonlinear constraints.

$$\text{(SCQ)} \Longrightarrow \text{(ACQ)} \Longrightarrow \text{(GCQ)}.$$

Local First-Order Optimality Conditions

- Convex optimization problems

$$(P)\begin{cases}\min\ f(\mathbf{x}),\\ g_i(\mathbf{x})\le 0,\ \text{for } i\in\mathcal{I}:=\{1,...,m\}.\end{cases}$$

The Lagrangian L to (P) is

$$L(\mathbf{x},\boldsymbol{\lambda}):=f(\mathbf{x})+\sum_{i=1}^{m}\lambda_i g_i(\mathbf{x})=f(\mathbf{x})+\langle\boldsymbol{\lambda},g(\mathbf{x})\rangle \text{ for } \mathbf{x}\in C \text{ and } \boldsymbol{\lambda}\in\mathbb{R}^m_+.$$

Definition 1. *A pair* $(\mathbf{x}^*,\boldsymbol{\lambda}^*)\in C\times R^m_+$ *is called a* saddle point *of* L *if and only if*

$$L(\mathbf{x}^*,\boldsymbol{\lambda})\le L(\mathbf{x}^*,\boldsymbol{\lambda}^*)\le L(\mathbf{x},\boldsymbol{\lambda}^*)$$

holds for all $\mathbf{x}\in C$ *and* $\boldsymbol{\lambda}\in\mathbb{R}^m_+$*, that is,* $\mathbf{x}^*$ *minimizes* $L(\cdot,\boldsymbol{\lambda}^*)$ *and* $\boldsymbol{\lambda}^*$ *maximizes* $L(\mathbf{x}^*,\cdot)$*.*

Local First-Order Optimality Conditions

- Convex optimization problems

Lemma 2. *If* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *is a saddle point of* L*, then it holds that:*

- $\mathbf{x}^*$ *is a global minimizer of (P).*
- $L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = f(\mathbf{x}^*)$.
- $\lambda_i^* g_i(\mathbf{x}^*) = 0$ *for all* $i \in \mathcal{I}$.

Local First-Order Optimality Conditions

- Convex optimization problems

We assume now that C is open and convex and the functions $f, g_i : C \to \mathbb{R}$ are continuously differentiable and convex for $i \in \mathcal{I}$. In this case we write more precisely (CP) instead of (P).

Theorem 1. *If the Slater constraint qualification holds and $\mathbf{x}^*$ is a minimizer of (CP), then there exists a vector $\boldsymbol{\lambda}^* \in \mathbb{R}^m_+$ such that $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is a saddle point of L.*

Proof. Taking into account our observations from *Constraint Qualifications*, KKT Theorem gives that there exists a $\boldsymbol{\lambda}^* \in \mathbb{R}^m_+$ such that

$$0 = L_x(\mathbf{x}^*, \boldsymbol{\lambda}^*) \text{ and } \langle \boldsymbol{\lambda}^*, g(\mathbf{x}^*) \rangle = 0.$$

With that we get for $\mathbf{x} \in C$,

$$L(\mathbf{x}, \boldsymbol{\lambda}^*) - L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \geq L_x(\mathbf{x}^*, \boldsymbol{\lambda}^*)(\mathbf{x} - \mathbf{x}^*) = 0$$

and

$$L(\mathbf{x}^*, \boldsymbol{\lambda}^*) - L(\mathbf{x}^*, \boldsymbol{\lambda}) = -\langle \underbrace{\boldsymbol{\lambda}}_{\geq 0}, \underbrace{g(\mathbf{x}^*)}_{\leq 0} \rangle \geq 0.$$

Hence, $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is a saddle point of L. □

Local First-Order Optimality Conditions

- Convex optimization problems

The Slater constraint qualification is essential.
Example:

$$(P)\begin{cases}\min\limits_{x} \ f(x) := -x, \\ g(x) := x^2 \le 0.\end{cases}$$

The only feasible point is $x^* = 0$ with value $f(0) = 0$. So 0 minimizes $f(x)$ subject to $g(x) \le 0$.

$L(x, \lambda) := -x + \lambda x^2$ for $\lambda \ge 0, x \in \mathbb{R}$. There is no $\lambda^* \in [0, \infty)$ such that (x^*, λ^*) is a saddle point of L.

Local First-Order Optimality Conditions

- Convex optimization problems

Constraint qualifications are not needed for a *sufficient* condition for general *convex optimization problems*.

Suppose that $f, g_i, h_j : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable functions with f and g_i convex and h_j (affinely) linear ($i \in \mathcal{I}, j \in \mathcal{E}$), and consider the following convex optimization problem:

$$(CP) \begin{cases} \min\limits_{\mathbf{x}} \ f(\mathbf{x}), \\ g_i(\mathbf{x}) \leq 0, \ \text{for } i \in \mathcal{I} \\ h_j(\mathbf{x}) = 0, \ \text{for } j \in \mathcal{E}. \end{cases}$$

We will show that *for this special kind of problem every KKT point already gives a (global) minimum.*

Local First-Order Optimality Conditions

- Convex optimization problems

Theorem 2. *Suppose* $\mathbf{x}_0 \in \mathcal{F}$ *and there exist* $\boldsymbol{\lambda} \in \mathbb{R}_+^m$ *and* $\boldsymbol{\mu} \in \mathbb{R}^p$ *such that*

$$\nabla f(\mathbf{x}_0) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x_0) + \sum_{j=1}^{p} \mu_i \nabla h_j(x_0) = \mathbf{0} \ \textit{and}$$

$$\lambda_i g_i(\mathbf{x}_0) = 0 \ \textit{for} \ i = 1, ..., m,$$

then (CP) attains its global minimum at $\mathbf{x}_0$.

Proof. By convexity, we get for $\mathbf{x} \in \mathcal{F}$:

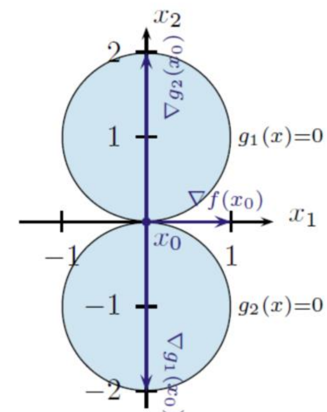
$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{x}_0) &\underset{f \text{ convex}}{\geq} f'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \\ &= -\sum_{i=1}^{m} \lambda_i g_i'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) - \sum_{j=1}^{p} \mu_j \underbrace{h_j'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)}_{=h_j(\mathbf{x}) - h_j(\mathbf{x}_0) = 0} \\ &\underset{g_i \text{ convex}}{\geq} -\sum_{i=1}^{m} \lambda_i (g_i(\mathbf{x}) - g_i(\mathbf{x}_0)) = -\sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \geq 0. \end{aligned}$$

Local First-Order Optimality Conditions

- Convex optimization problems

Even if we have convex problems the KKT conditions are not necessary for minimal points.
Example:

$$(P) \begin{cases} \min\limits_{\mathbf{x}} \; f(\mathbf{x}) := x_1, \\ g_1(\mathbf{x}) := x_1^2 + (x_2 - 1)^2 - 1 \leq 0, \\ g_2(\mathbf{x}) := x_1^2 + (x_2 + 1)^2 - 1 \leq 0. \end{cases}$$



Obviously, only the point $\mathbf{x}_0 := (0,0)^\top$ is feasible. Hence, $\mathbf{x}_0$ is the (global) minimal point. Since $\nabla f(\mathbf{x}_0) = (1,0)^\top$, $\nabla g_1(\mathbf{x}_0) = (0,-2)^\top$ and $\nabla g_2(\mathbf{x}_0) = (0,2)^\top$, the gradient condition of the KKT conditions is not met.
Note that the Slater condition is not fulfilled.

Local First-Order Optimality Conditions

- Relationship among the conditions

$$
\begin{tikzcd}
\mathbf{x}_0 \text{ is a local minimizer} \arrow[r, shift right, Rightarrow, "\text{CQ}"'] \arrow[dr, Rightarrow] & \mathcal{C}_l(\mathbf{x}_0) \cap \mathcal{C}_{dd}(\mathbf{x}_0) = \emptyset \arrow[l, shift right, Rightarrow, "\text{CP}"'] \arrow[r, Leftrightarrow] \arrow[d, Rightarrow] & \text{KKT conditions} \\
& \mathcal{C}_t(\mathbf{x}_0) \cap \mathcal{C}_{dd}(\mathbf{x}_0) = \emptyset \arrow[d, Rightarrow] & \\
& \mathcal{C}_{fd}(\mathbf{x}_0) \cap \mathcal{C}_{dd}(\mathbf{x}_0) = \emptyset &
\end{tikzcd}
$$

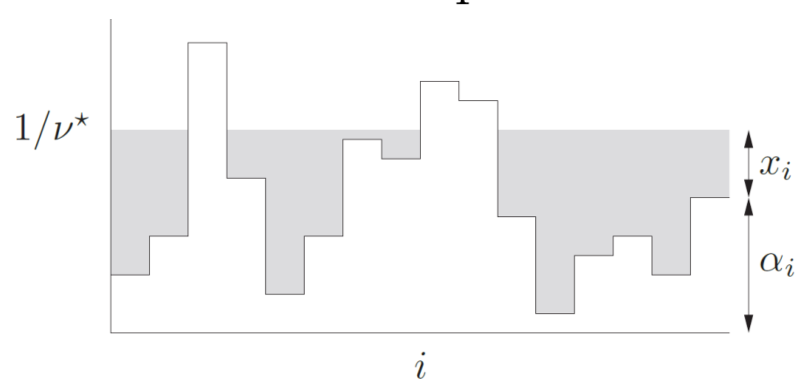
Optimality conditions

- KKT optimality conditions for convex problems

If a convex optimization problem with differentiable objective and constraint functions satisfies Slater's condition, then the KKT conditions provide *necessary and sufficient conditions* for optimality: Slater's condition implies that the optimal duality gap is zero and the dual optimum is attained, so $\mathbf{x}$ is optimal iff there are $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ that, together with $\mathbf{x}$, satisfy the KKT conditions.

The KKT conditions play an important role in optimization. In a few special cases it is possible to solve the KKT conditions (and therefore, the optimization problem) analytically. More generally, many algorithms for convex optimization are conceived as, or can be interpreted as, methods for solving the KKT conditions.

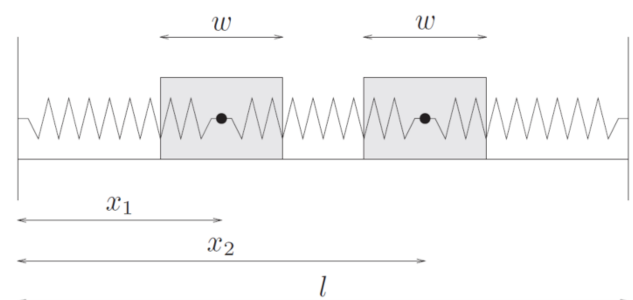
Examples: Equality constrained convex quadratic minimization, Water-filling



Optimality conditions

- Mechanics interpretation of KKT

The KKT conditions can be given a nice interpretation in mechanics (which indeed, was one of Lagrange's primary motivations). We illustrate the idea with a simple system shown in the figure. The position of the blocks are given by $\mathbf{x} \in \mathbb{R}^2$ where x_1 is the displacement of the (middle of the) left block, and x_2 is the displacement of the right block. The left wall is at position 0, and the right wall is at position l.



The potential energy in the springs is given by

$$f_0(x_1, x_2) = \frac{1}{2}k_1x_1^2 + \frac{1}{2}k_2(x_2 - x_1)^2 + \frac{1}{2}k_3(l - x_2)^2,$$

where $k_i > 0$ are the stiffness constants of the three springs. The equilibrium position $\mathbf{x}^*$ is the position that minimizes the potential energy subject to

$$w/2 - x_1 \leq 0, \quad w + x_1 - x_2 \leq 0, \quad w/2 - l + x_2 \leq 0. \tag{19}$$

Optimality conditions

- Mechanics interpretation of KKT

These constraints are called *kinematic constraints*, and express the fact that the blocks have width $w > 0$, and cannot penetrate each other or the walls. The equilibrium position is therefore given by the solution of the optimization problem

$$\begin{array}{ll} \text{minimize} & (1/2)(k_1 x_1^2 + k_2(x_2 - x_1)^2 + k_3(l - x_2)^2) \\ \text{subject to} & w/2 - x_1 \leq 0 \\ & w + x_1 - x_2 \leq 0 \\ & w/2 - l + x_2 \leq 0, \end{array} \tag{20}$$

which is a QP.

Optimality conditions

- Mechanics interpretation of KKT

With λ_1, λ_2, λ_3 as Lagrange multipliers, the KKT conditions for this problem consist of the kinematic constraints (19), the nonnegativity constraints $\lambda_i \geq 0$, the complementary slackness conditions

$$\lambda_1(w/2 - x_1) = 0, \quad \lambda_2(w - x_2 + x_1) = 0, \quad \lambda_3(w/2 - l + x_2) = 0, \tag{21}$$

and the zero gradient condition

$$\begin{bmatrix} k_1 x_1 - k_2(x_2 - x_1) \\ k_2(x_2 - x_1) - k_3(l - x_2) \end{bmatrix} + \lambda_1 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \lambda_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0. \tag{22}$$

The equation (21) can be interpreted as the force balance equations for the two blocks, provided we interpret the Lagrange multipliers as *contact forces* that act between the walls and blocks.

Optimality conditions

- Mechanics interpretation of KKT

The first equation states that the sum of the forces on the first block is zero: The term $-k_1x_1$ is the force exerted on the left block by the left spring, the term $k_2(x_2 - x_1)$ is the force exerted by the middle spring, λ_1 is the force exerted by the left wall, and $-\lambda_2$ is the force exerted by the right block. The contact forces must point away from the contact surface (as expressed by the constraints $\lambda_1 \geq 0$ and $-\lambda_2 \leq 0$), and are nonzero only when there is contact (as expressed by the first two complementary slackness conditions (21)). In a similar way, the second equation in (22) is the force balance for the second block, and the last condition in (21) states that λ_3 is zero unless the right block touches the wall.

In this example, the potential energy and kinematic constraint functions are convex, and (the refined form of) Slater's constraint qualification holds provided $2w \leq l$, *i.e.*, there is enough room between the walls to fit the two blocks, so we can conclude that the energy formulation of the equilibrium given by (20), gives the same result as the force balance formulation, given by the KKT conditions.

Duality

- The Lagrange dual function - The Lagrangian

We consider an optimization problem in the standard form:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & f_0(\mathbf{x}), \\ \text{s. t.} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \cdots, m, \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \cdots, p. \end{aligned} \tag{1}$$

We assume its domain $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\ f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\ h_i$ is nonempty, and denote the optimal value of (1) by p^*.

Duality

- The Lagrange dual function - The Lagrangian

We define the *Lagrangian* $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ associated with the problem (1) as

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}),$$

with $\mathbf{dom}\, L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$. We refer to λ_i as the *Lagrange multiplier* associated with the ith inequality constraint $f_i(x) \leq 0$; similarly we refer to ν_i as the *Lagrange multiplier* associated with the ith equality constraint $h_i(x) = 0$. The vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ are called the *dual variables or Lagrange multiplier vectors* associated with the problem (1).

Duality

- The Lagrange dual function - The Lagrange dual function

We define the *Lagrange dual function* (or just *dual function*) $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ as the minimum value of the Lagrangian over $\mathbf{x}$: for $\boldsymbol{\lambda} \in \mathbb{R}^m$, $\boldsymbol{\nu} \in \mathbb{R}^p$,

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \right).$$

When the Lagrangian is unbounded below in $\mathbf{x}$, the dual function takes on the value $-\infty$. Since the dual function is the pointwise infimum of a family of affine functions of $(\boldsymbol{\lambda}, \boldsymbol{\nu})$, it is concave, even when the problem (1) is not convex.

Duality

- The Lagrange dual function - Lower bounds on optimal value

For any $\boldsymbol{\lambda} \geq \mathbf{0}$ and any $\boldsymbol{\nu}$ we have an important relationship:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*. \tag{2}$$

Proof. Suppose $\tilde{\mathbf{x}}$ is a feasible point for the problem (1), *i.e.*, $f_i(\tilde{\mathbf{x}}) \leq 0$ and $h_i(\tilde{\mathbf{x}}) = 0$, and $\boldsymbol{\lambda} \geq 0$. Then we have

$$\sum_{i=1}^{m} \lambda_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{\mathbf{x}}) \leq 0.$$

Therefore

$$L(\tilde{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{\mathbf{x}}) \leq f_0(\tilde{\mathbf{x}}).$$

Hence

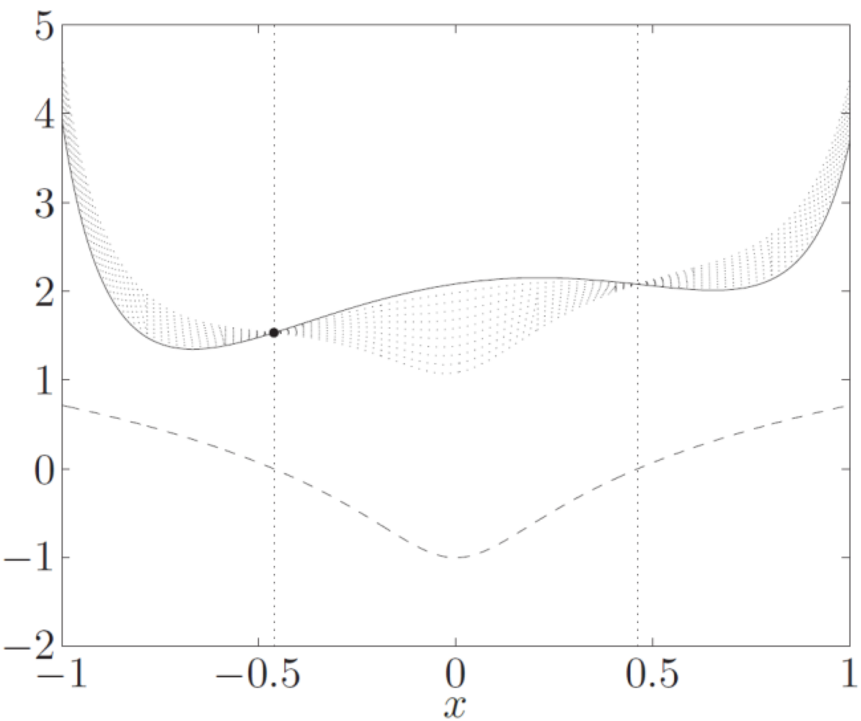
$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq L(\tilde{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f_0(\tilde{\mathbf{x}}).$$

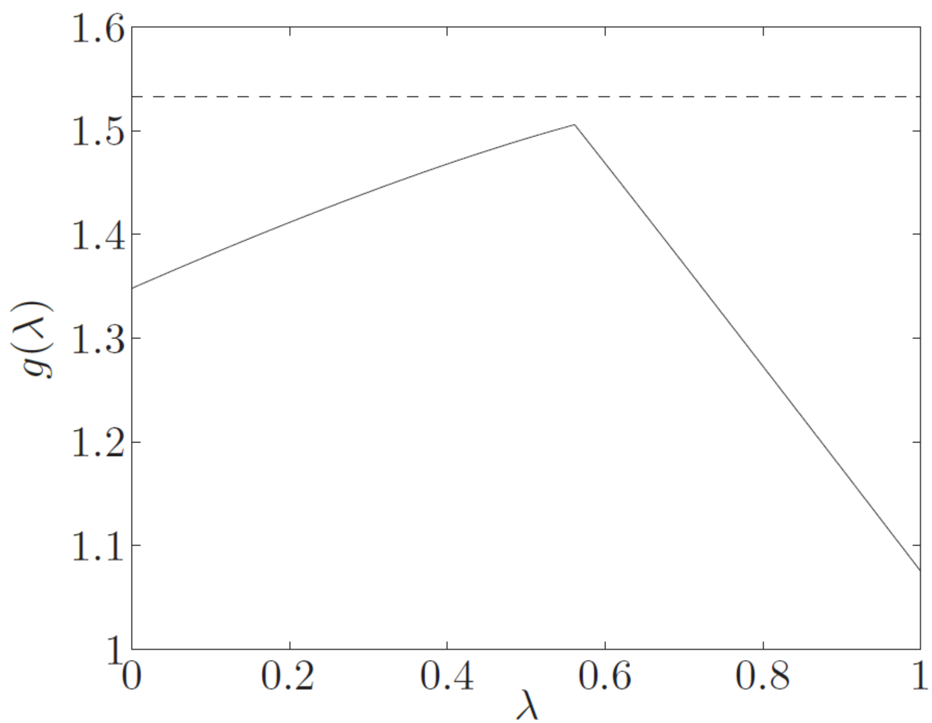
Since $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f_0(\tilde{\mathbf{x}})$ holds for every feasible point $\tilde{\mathbf{x}}$, the inequality (2) follows.

Duality

- The Lagrange dual function - Lower bounds on optimal value

The *domain* of dual function g is $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ such that $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty$. We refer to a pair $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ with $\boldsymbol{\lambda} \geq \mathbf{0}$ and $(\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \mathbf{dom}\ g$ as *dual feasible*.





Duality

- The Lagrange dual function - Linear approximation interpretation

The Lagrangian and lower bound property can be given a simple interpretation, based on a linear approximation of the indicator functions of the sets $\{0\}$ and $-\mathbb{R}_+$.

We first rewrite the original problem (1) as an unconstrained problem,

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) + \sum_{i=1}^{p} I_0(h_i(x)), \tag{3}$$

where $I_- : \mathbb{R} \to \mathbb{R}$ is the indicator function for the nonpositive reals,

$$I_-(u) = \begin{cases} 0, & u \leq 0, \\ \infty, & u > 0, \end{cases}$$

and similarly, I_0 is the indicator function of $\{0\}$.

Duality

- The Lagrange dual function - Linear approximation interpretation

Now suppose in the formulation (3) we replace the function $I_{-}(u)$ with the linear function $\lambda_i u$, where $\lambda_i \geq 0$, and the function $I_0(u)$ with $\nu_i u$. The objective becomes the Lagrangian function $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$, and the dual function value $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is the optimal value of the problem

$$\min_{\mathbf{x}} \quad L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}). \tag{4}$$

In this formulation, we use a linear or "soft" displeasure function in place of I_{-} and I_0. For an inequality constraint, our displeasure is zero when $f_i(\mathbf{x}) = 0$, and is positive when $f_i(x) > 0$ (assuming $\lambda_i > 0$); our displeasure grows as the constraint becomes "more violated". Unlike the original formulation, in which any nonpositive value of $f_i(\mathbf{x})$ is acceptable, in the soft formulation we actually derive pleasure from constraints that have margin, *i.e.*, from $f_i(\mathbf{x}) < 0$.

Duality

- The Lagrange dual function - Linear approximation interpretation

Clearly the approximation of the indicator function $I_{-}(u)$ with a linear function $\lambda_i u$ is rather poor. But the linear function is at least an *underestimator* of the indicator function. Since $\lambda_i u \leq I_{-}(u)$ and $\nu_i u \leq I_0(u)$ for all u, we see immediately that the dual function yields a lower bound on the optimal value of the original problem.

Examples: Least-squares solution of linear equations, Standard form LP, Two-way partitioning problem,

Duality

- The Lagrange dual function - The Lagrange dual function and conjugate functions

The conjugate function and Lagrange dual function are closely related. To see one simple connection, consider the problem

$$\begin{aligned} &\min\ f(\mathbf{x}), \\ &s.t.\ \mathbf{x} = \mathbf{0}. \end{aligned}$$

This problem has Lagrangian $L(\mathbf{x}, \boldsymbol{\nu}) = f(\mathbf{x}) + \boldsymbol{\nu}^T\mathbf{x}$, and dual function

$$g(\boldsymbol{\nu}) = \inf_{\mathbf{x}} \left(f(\mathbf{x}) + \boldsymbol{\nu}^T\mathbf{x}\right) = -\sup_{\mathbf{x}} \left((-\boldsymbol{\nu})^T\mathbf{x} - f(\mathbf{x})\right) = -f^*(-\boldsymbol{\nu}).$$

Duality

- The Lagrange dual function - The Lagrange dual function and conjugate functions

More generally (and more usefully), consider an optimization problem with linear inequality and equality constraints,

$$\begin{aligned}
&\min_{\mathbf{x}} \ f_0(\mathbf{x}), \\
&s.t. \ \mathbf{A}\mathbf{x} \preceq \mathbf{b}, \\
&\qquad \mathbf{C}\mathbf{x} = \mathbf{d}.
\end{aligned}$$

Using the conjugate of f_0 we can write its dual function as

$$\begin{aligned}
g(\boldsymbol{\lambda}, \boldsymbol{\nu}) &= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) + \boldsymbol{\nu}^T(\mathbf{C}\mathbf{x} - \mathbf{d}) \right) \\
&= -\mathbf{b}^T\boldsymbol{\lambda} - \mathbf{d}^T\boldsymbol{\nu} + \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + (\mathbf{A}^T\boldsymbol{\lambda} + \mathbf{C}^T\boldsymbol{\nu})^T\mathbf{x} \right) \\
&= -\mathbf{b}^T\boldsymbol{\lambda} - \mathbf{d}^T\boldsymbol{\nu} - f_0^*(-\mathbf{A}^T\boldsymbol{\lambda} - \mathbf{C}^T\boldsymbol{\nu}).
\end{aligned}$$

The domain of g follows from the domain of f_0^*:

$$\mathbf{dom}\ g = \{(\boldsymbol{\lambda}, \boldsymbol{\nu}) |\ \boldsymbol{\lambda} \geq \mathbf{0}, -\mathbf{A}^T\boldsymbol{\lambda} - \mathbf{C}^T\boldsymbol{\nu} \in \mathbf{dom}\ f_0^*\}.$$

Duality

- The Lagrange dual function - The Lagrange dual function and conjugate functions

Examples: Equality constrained norm minimization, Entropy maximization, Minimum volume covering ellipsoid.