Chapter 6: Optimality Condition

Chao Wang

SUSTech

Outline

1. Necessary Conditions
2. Sufficient Conditions
3. Conditions for Convex Problems
4. Lagrangian, Duality, & Complementary Slackness
5. Karush-Kuhn-Tucker (KKT) Conditions

Necessary Conditions

Optimization problem

- The general form of optimization:

$$\begin{array}{rl} \text{Min} & f(x) \\ \text{Subject to} & x \in \Omega \end{array}$$

- Suppose $x \in \mathbb{R}^n, \Omega$ is called the feasible set.
- If $\Omega = \mathbb{R}^n$, then the problem is called unconstrained.
- Otherwise, the problem is called constrained.
- We can write any constrained problem in the unconstrained form

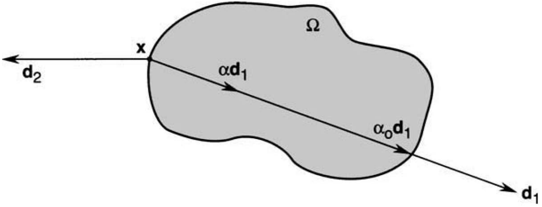
$$\min f(x) + \iota_\Omega(x),$$

where the indicator function

$$\iota_\Omega(x) = \begin{cases} 0, x \in \Omega, \\ \infty, x \notin \Omega. \end{cases}$$

Feasible direction

- A vector $d \in \mathbb{R}^n$ is a feasible direction at $x \in \Omega$ if $d \neq 0$ and $x + \alpha d \in \Omega$ for some small $\alpha > 0$. (It is possible that d is an infeasible step, that is, $x + d \notin \Omega$. But if there is some room in Ω to move from x toward d, then d is a feasible direction.)
- If $\Omega = \mathbb{R}^n$ or x lies in the interior of Ω, then any $d \in \mathbb{R}^n \backslash \{0\}$ is a feasible direction
- Feasible directions are introduced to establish optimality conditions, especially for points on the boundary of a constrained problem



d_1 is feasible, d_2 is infeasible

First-order necessary condition

$\phi(a+h) = \phi(a) + \phi'(a)h + \frac{\phi''(a)}{2}h^2 + \cdots + \frac{\phi^{(m)}(a)}{m!}h^m + o(h^m).$

Theorem

(First-Order Necessary Condition (FONC)). *Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real-value function on Ω. If x^* is a local minimizer of f over Ω, then for any feasible direction d at x^*, we have*

$$d^T \nabla f(x^*) \geq 0$$

Proof: Let d by any feasible direction. First-order Taylor expansion:

$$f(x^* + \alpha d) = f(x^*) + \underbrace{\alpha d^T \nabla f(x^*)}_{<0} + o(\alpha).$$

反证.

$o(\alpha)$ 可忽略

If $d^T \nabla f(x^*) < 0$, which does not depend on α, then $f(x^* + \alpha d) < f(x^*)$ for all sufficiently small $\alpha > 0$ (that is, all $\alpha \in (0, \bar{\alpha})$ for some $\bar{\alpha} > 0$). This is a contradiction since x^* is a local minimizer.

与x^*区域最小值相违.

First-order necessary condition

Corollary

(Interior Case) *Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real-value function on Ω. If x^* is a local minimizer of f over Ω and if x^* is an interior point, then $\nabla f(x^*) = 0$.*

any d s.t. $d^T \nabla f(x^*) \geq 0$.

d 都为 feasible direction.

Proof: Since any $d \in \mathbb{R}^n \backslash \{0\}$ is a feasible direction, we can set $d = -\nabla f(x^*)$. We have $d^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 \geq 0$. Since $\|\nabla f(x^*)\|^2 \geq 0$, we have $\|\nabla f(x^*)\|^2 = 0$ and thus $\nabla f(x^*) = 0$.

Comment: This condition also reduces the problem

$$\min f(x)$$

to solving the equation

$$\nabla f(x^*) = 0$$

Second-order necessary condition

In FONC, there are two possibilities

- $d^T \nabla f(x^*) > 0$;
- $d^T \nabla f(x^*) = 0$.

In the first case, $f(x^* + \alpha d) > f(x^*)$ for all sufficiently small $\alpha > 0$. In the second case, the vanishing $d^T \nabla f(x^*)$ allows us to check higher-order derivatives.

Theorem

(Second-Order Necessary Condition (SONC)) *Let $\Omega \subseteq \mathbb{R}^n$, $f \in \mathcal{C}^2$* ① *a function on Ω, x^* a local minimizer of f over Ω, and d a* ∀ 满足 *feasible direction at x^*. If $d^T \nabla f(x^*) = 0$,* ②. *then*

$$d^T F(x^*) d \geq 0$$

where F is the Hessian of f.

反证.

Proof: Assume that $\exists$ a feasible direction d with $d^T \nabla f(x^*) = 0$ and $d^T F(x^*) d < 0$. By 2 nd-order Taylor expansion (with a vanishing 1 st order term), we have

为0,

$$f(x^* + \alpha d) = f(x^*) + \frac{d^T F(x^*) d}{2} \alpha^2 + o\left(\alpha^2\right),$$

where by our assumption $d^T F(x^*) d < 0$. Hence, for all sufficiently small $\alpha > 0$, we have $f(x^* + \alpha d) < f(x^*)$, which contradicts that x^* is a local minimizer.

Second-order necessary condition

Corollary

(Interior Case) *Let x^* be a interior point of $\Omega \subset \mathbb{R}^n$. If x^* is a local minimizer of $f : \Omega \to \mathbb{R}^n, f \in \mathcal{C}^2$, then*

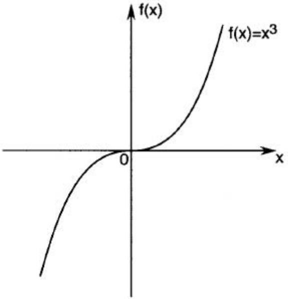
$$\nabla f(x^*) = 0,$$

and $F(x^*)$ is positive semidefinite $(F(x^*) \geq 0)$; that is, for all $d \in \mathbb{R}^n$,

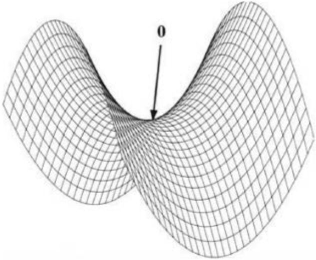
$$d^T F(x^*) d \geq 0.$$

The necessary conditions are not sufficient

Counter examples



$f(x) = x^3$, $f'(x) = 3x^2$, $f''(x) = 6x$



$f(x) = x_1^2 - x_2^2$

$\mathbf{0}$ is a saddle point: $\nabla f(\mathbf{0}) = \mathbf{0}$ but neither a local minimizer nor maximizer

By SONC, $\mathbf{0}$ is not a local minimizer!

Sufficient Conditions

Second-order sufficient condition

Theorem

(Second-Order Sufficient Condition (SOSC), Interior point.) *Let $f \in \mathcal{C}^2$ be defined on a region in which x^* is an interior point. Suppose that*

1. $\nabla f(x^*) = 0$*;*

2. $F(x^*) > 0$*.*

Then, x^* is a strict local minimizer of f.

Comments:

- part 2 states $F(x^*)$ is positive definite: $x^T F(x^*) x > 0$ for $x \neq 0$.
- the condition is not necessary for strict local minimizer.

Proof: For any $d \neq 0$ and $\|d\| = 1$, we have $d^T F(x^*) d \geq \lambda_{\min}(F(x^*)) > 0$. Use the 2nd order Taylor expansion

$$f(x^* + \alpha d) = f(x^*) + \frac{\alpha^2}{2} d^T F(x^*) d + o\left(\alpha^2\right) \geq f(x^*) + \frac{\alpha^2}{2} \lambda_{\min}(F(x^*)) + o\left(\alpha^2\right)$$

Then, $\exists \bar{\alpha} > 0$, regardless of d, such that $f(x^* + \alpha d) > f(x^*), \alpha \in (0, \bar{\alpha})$.

Conditions for Convex Problems

Optimality conditions for convex problem

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function, and let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Consider the problem

$$\begin{array}{ll} \textit{minimize} & f(x) \\ \textit{subject to} & x \in C. \end{array}$$

A vector x^* is optimal for this problem if and only if $x^* \in C$ and

$$\nabla f(x^*)^T (z - x^*) \geq 0 \textit{ for all } z \in C.$$

Proof. For the sake of simplicity, we prove the result assuming that f is continuously differentiable. Let x^* be optimal. Suppose that for some $\hat{z} \in C$ we have

$$\nabla f(x^*)^T (\hat{z} - x^*) < 0.$$

Optimality conditions for convex problem

Since f is continuously differentiable, by the first-order Taylor expansion, we have for all sufficiently small $\alpha > 0$,

$$f\left(x^*+\alpha\left(\hat{z}-x^*\right)\right)=f\left(x^*\right)+\alpha \nabla f\left(x^*\right)^T\left(\hat{z}-x^*\right)+o(\alpha)<f\left(x^*\right),$$

with $x^* \in C$ and $\hat{z} \in C$. By the convexity of C, we have $x^*+\alpha\left(\hat{z}-x^*\right) \in C$. Thus, this vector is feasible and has a smaller objective value than the optimal point x^*, which is a contradiction. Hence, we must have $\nabla f\left(x^*\right)^T\left(z-x^*\right) \geq 0$ for all $z \in C$. Suppose now that $x^* \in C$ and

$$\nabla f\left(x^*\right)^T\left(z-x^*\right) \geq 0 \text{ for all } z \in C.$$

By convexity of f, we have

$$f\left(x^*\right)+\nabla f\left(x^*\right)^T\left(z-x^*\right) \leq f(z) \text{ for all } z \in C,$$

implying that

$$\nabla f\left(x^*\right)^T\left(z-x^*\right) \leq f(z)-f\left(x^*\right).$$

This and Eq. (2.10) further imply that

$$0 \leq f(z)-f\left(x^*\right) \text{ for all } z \in C.$$

Since $x^* \in C$, it follows that x^* is optimal.

Optimality conditions for convex problem

- We next discuss several implications of the theorem, by considering some special choices for the set C. Let C be the entire space, i.e., $C = \mathbb{R}^n$. The condition

$$\nabla f\left(x^*\right)^T \left(z - x^*\right) \geq 0 \text{ for all } z \in C$$

reduces to

$$\nabla f\left(x^*\right)^T d \geq 0 \text{ for all } d \in \mathbb{R}^n$$

- this is equivalent to

$$\nabla f\left(x^*\right) = 0.$$

Thus, a vector x^* is a minimum of f over $\mathbb{R}^n$ if and only if $\nabla f\left(x^*\right) = 0$.

Lagrangian, Duality, & Complementary Slackness

Lagrangian

- Standard form problem (not necessarily convex)

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

  variable $x \in \mathbf{R}^n$, domain $\mathcal{D}$, optimal value $p^\star$

- Lagrangian: $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$, with dom $L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$,

$$L(x, \lambda, v) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} v_i h_i(x)$$

  - weighted sum of objective and constraint functions
  - λ_i is Lagrange multiplier associated with $f_i(x) \leq 0$
  - v_i is Lagrange multiplier associated with $h_i(x) = 0$

Lagrangian dual function

- Lagrange dual function: $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

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

  - a concave function of λ, v
  - can be $-\infty$ for some λ, v; this defines the domain of g

- Lower bound property: if $\lambda \succeq 0$, then $g(\lambda, v) \le p^\star$
  proof: if x is feasible and $\lambda \succeq 0$, then

$$f_0(x) \ge L(x, \lambda, v) \ge \inf_{\tilde{x} \in \mathcal{D}} L(\tilde{x}, \lambda, v) = g(\lambda, v)$$

  minimizing over all feasible x gives $p^\star \ge g(\lambda, v)$

Connect the dual function in LP

- LP problem:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

- its Lagrangian is

$$\begin{aligned} L(x, \lambda, v) &= c^T x + v^T (Ax - b) - \lambda^T x \\ &= -b^T v + \left(c + A^T v - \lambda\right)^T x \end{aligned}$$

- L is affine in x, hence

$$g(\lambda, v) = \inf_x L(x, \lambda, v) = \begin{cases} -b^T v & A^T v - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

- g is linear on affine domain $\operatorname{dom} g = \left\{(\lambda, v) | A^T v - \lambda + c = 0\right\}$, hence concave
- Lower bound property: $p^\star \geq -b^T v$ if $A^T v + c \geq 0$

Other example

- Least norm solution of linear equations:

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b \end{array}$$

- Lagrangian is

$$L(x, v) = x^T x + v^T (Ax - b)$$

- to minimize L over x, set gradient equal to zero:

$$\nabla_x L(x, v) = 2x + A^T v = 0 \Longrightarrow x = -\frac{1}{2} A^T v$$

- plug in in L to obtain g :

$$g(v) = L\left(-\frac{1}{2} A^T v, v\right) = -\frac{1}{4} v^T A A^T v - b^T v$$

  a concave function of v

- Lower bound property: $p^\star \geq -\frac{1}{4} v^T A A^T v - b^T v$ for all v

Other example

- equality constrained norm minimization:

$$\begin{array}{ll} \text{minimize} & \|x\| \\ \text{subject to} & Ax = b \end{array}$$

- $\|\cdot\|$ is any norm; dual norm is defined as

$$\|v\|_* = \sup_{\|u\| \leq 1} u^T v$$

- define Lagrangian $L(x, v) = \|x\| + v^T(b - Ax)$
- dual function (tutorial):

$$g(v) = \inf_x \left(\|x\| - v^T Ax + b^T v \right)$$
$$= \begin{cases} b^T v & \left\|A^T v\right\|_* \leq 1 \\ -\infty & \text{otherwise} \end{cases}$$

- Lower bound property: $p^\star \geq b^T v$ if $\left\|A^T v\right\|_* \leq 1$

$$\min \|x\| - y^Tx = \begin{cases} 0 & \|y\|_* \le 1 \\ -\infty & \text{otherwise} \end{cases}$$

$$\|v\|_* = \sup_{\|u\| \le 1} u^Tv$$

p.f.:

Case 1.: $\|y\|_* \le 1$.

let $u = \dfrac{x}{\|x\|}$ $\quad \dfrac{x^T}{\|x\|} y \le \|y\|_* \Rightarrow \|x\| \underbrace{\|y\|_*}_{\le 1} \ge x^Ty$

$\Downarrow$

$\|x\| - y^Tx \ge 0. \quad \forall x$

$\Rightarrow$ let $x = 0$ $\quad \|x\| - y^Tx = 0$

Case II: $\|y\|_* > 1$

$\exists \tilde{x}$ $\quad \|\tilde{x}\| \le 1$ $\quad y^T\tilde{x} = \|y\|_*$

$\Rightarrow \|\tilde{x}\| - \|y\|_* < 0.$

$\forall t > 0$ $\quad x = t\tilde{x}$

$\|x\| - y^Tx = t(\|\tilde{x}\| - \|y\|_*)$

Lagrange dual and conjugate function

- Conjugate function of f: $f^*(y) = \sup_x \left(y^T x - f(x)\right)$
- consider

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax \leq b \\ & Cx = d \end{array}$$

- Its dual function is

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

- simplifies derivation of dual if conjugate of f_0 is known

Example: entropy maximization

$$f_0(x) = \sum_{i=1}^n x_i \log x_i, \; f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

The dual problem

Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda, v) \\ \text{subject to} & \lambda \geq 0 \end{array}$$

- finds best lower bound on $p^\star$, obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted by $d^\star$
- often simplified by making implicit constraint $(\lambda, v) \in \operatorname{dom} g$ explicit
- λ, v are dual feasible if $\lambda \geq 0, (\lambda, v) \in \operatorname{dom} g$
- $d^\star = -\infty$ if problem is infeasible; $d^\star = +\infty$ if unbounded above

Example: standard form LP and its dual

$$\begin{array}{llll} \text{minimize} & c^T x & \text{maximize} & -b^T v \\ \text{subject to} & Ax = b & \text{subject to} & A^T v + c \geq 0 \\ & x \geq 0 & & \end{array}$$

Weak and strong duality

- Weak duality: $d^\star \leq p^\star$
  - always holds (for convex and nonconvex problems)
  - can be used to find nontrivial lower bounds for difficult problems
- Strong duality: $d^\star = p^\star$
  - does not hold in general
  - (usually) holds for convex problems
  - sufficient conditions that guarantee strong duality in convex problems are called constraint qualifications

Inequality from LP

- Primal problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \end{array}$$

- Dual function

  $$g(\lambda) = \inf_x \left(\left(c + A^T\lambda\right)^T x - b^T\lambda \right) = \begin{cases} -b^T\lambda & A^T\lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

  Dual problem

$$\begin{array}{ll} \text{maximize} & -b^T\lambda \\ \text{subject to} & A^T\lambda + c = 0 \\ & \lambda \geq 0 \end{array}$$

- Primal problem (assume $P \in \mathbf{S}^n_{++}$)

$$\begin{array}{ll} \text{minimize} & x^T P x \\ \text{subject to} & Ax \le b \end{array}$$

- Dual function

$2P^Tx + \lambda^T A = 0$

$2P^T x = -\lambda^T A$

$$g(\lambda) = \inf_x \left(x^T P x + \lambda^T (Ax - b) \right) = -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

$2x = -\lambda^T P^{-1} A$

$x = -\frac{1}{2}$

$\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - \frac{1}{2} \lambda^T A P^{-1} A \lambda$

- Dual problem

$$\begin{array}{ll} \text{maximize} & -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ \text{subject to} & \lambda \ge 0 \end{array}$$

  - in fact, $p^\star = d^\star$ always

Complementary slackness

assume x satisfies the primal constraints and $\lambda \geq 0$

$$
\begin{aligned}
g(\lambda, v) &= \inf_{\tilde{x} \in \mathcal{D}} \left(f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} v_i^{\star} h_i(\tilde{x}) \right) \\
&\leq f_0(x) + \underbrace{\sum_{i=1}^{m} \lambda_i f_i(x)}_{=0} + \underbrace{\sum_{i=1}^{p} v_i h_i(x)}_{=0} \\
(=) &\leq f_0(x)
\end{aligned}
$$

equality $f_0(x) = g(\lambda, v)$ holds if and only if the two inequalities hold with equality:

- first inequality: x minimizes $L(\tilde{x}, \lambda, v)$ over $\tilde{x} \in \mathcal{D}$
- 2nd inequality: $\lambda_i f_i(x) = 0$ for $i = 1, \ldots, m$, i.e.,

$$\lambda_i > 0 \Longrightarrow f_i(x) = 0, f_i(x) < 0 \Rightarrow \lambda_i = 0$$

this is known as complementary slackness

Karush-Kuhn-Tucker (KKT) Conditions

Optimality conditions

- if strong duality holds, then x is primal optimal and (λ, v) is dual optimal if:
  1. (primal feasibility) $f_i(x) \leq 0$ for $i = 1, \ldots, m$ and $h_i(x) = 0$ for $i = 1, \ldots, p$
  2. (dual feasibility) $\lambda \geq 0$
  3. $\lambda_i f_i(x) = 0$ for $i = 1, \ldots, m$
  4. x is a minimizer of $L(\cdot, \lambda, v)$

  $\Leftrightarrow$ optimal

  conversely, these four conditions imply optimality of $x, (\lambda, v)$, and strong duality if problem is convex and the functions f_i, h_i are differentiable, 4 can written as
- $4'$ (stationarity). the gradient of the Lagrangian with respect to x vanishes:

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

- conditions $1, 2, 3, 4'$ are known as Karush-Kuhn-Tucker (KKT) conditions

Convex problem

- strong duality: $p^\star = d^\star$
- if optimal value is finite, dual optimum is attained: there exist dual optimal λ, v

hence, if problem is convex

- x is optimal if and only if there exist λ, v such that $1 - 4$ are satisfied
- if functions are differentiable, condition 4 can be replaced with 4' (KKT conditions: 1-3,4')

Example: water-filling

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^{n} \log\left(x_i + \alpha_i\right) \\ \text{subject to} & x \geq 0 \\ & \mathbf{1}^T x = 1 \end{array}$$

- we assume that $\alpha_i > 0$
- Lagrangian is

$$L(\tilde{x}, \lambda, v) = -\sum_i \log\left(\tilde{x}_i + \alpha_i\right) - \lambda^T \tilde{x} + v\left(\mathbf{1}^T \tilde{x} - 1\right)$$

Optimality conditions: x is optimal iff there exist $\lambda \in \mathbb{R}^n, v \in \mathbb{R}$ such that

1. $x \geq 0, 1^T x = 1$
2. $\lambda \geq 0$
3. $\lambda_i x_i = 0$ for $i = 1, \ldots, n$
4. x minimizes Lagrangian: $\frac{1}{x_i + \alpha_i} + \lambda_i = v, \quad i = 1, \ldots, n$

(a). $f(x) \geq 0 \quad \forall x$ let $x=(0,0)$ $f(x)=0$ $(f(x) \geq 0)$.

(b). $f'(x)=0$ $f''(0) = \begin{pmatrix} 2 & -4 \\ -4 & 8 \end{pmatrix} \succeq 0$.

(c). $|f''(0)| = 0$ No

(d). Yes $(x_1, x_2) \neq (0,0)$, $f(x) > 0$.

2. $f'(x) = 1 - \frac{1}{x}$

Example: water-filling

Solution

- if $v \leq 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/v - \alpha_i$
- if $v \geq 1/\alpha_i$: $x_i = 0$ and $\lambda_i = v - 1/\alpha_i$
- two cases may be combined as

$$x_i = \max\left\{0, \frac{1}{v} - \alpha_i\right\}, \lambda_i = \max\left\{0, v - \frac{1}{\alpha_i}\right\}$$

- determine v from condition $\mathbf{1}^T x = 1$:

$$\sum_{i=1}^{n} \max\left\{0, \frac{1}{v} - \alpha_i\right\} = 1$$

Interpretation

- n patches; level of patch i is at height α_i
- flood area with unit amount of water
- resulting level is $1/v^{\star}$

Extra reading materials

More notes on KKT and duality:

- www.cs.cmu.edu/~./ggordon/10725-F12/scribes/10725_Lecture16.pdf
- www.ifp.illinois.edu/~angelia/L9_kktconditions.pdf
- www.stat.cmu.edu/~ryantibs/convexopt/lectures/kkt.pdf