Lagrange Duality

Ye Shi

ShanghaiTech University

Outline

Lagrangian

- Consider an optimization problem in standard form (not necessarily convex)

$$\begin{array}{lll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) & \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0 & i = 1, \cdots, m \\ & h_i(\boldsymbol{x}) = 0 & i = 1, \cdots, p \end{array}$$

  with variable $\boldsymbol{x} \in \mathbb{R}^n$, domain $\mathcal{D}$, and optimal value $p^\star$

- The *Lagrangian* is a function $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, with $\operatorname{dom} L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$, defined as

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

  where λ_i is the Lagrange multiplier associated with $f_i(\boldsymbol{x}) \leq 0$ and ν_i is the Lagrange multiplier associated with $h_i(\boldsymbol{x}) = 0$.

Outline

Lagrange Dual Function I

- The *Lagrange dual function* is defined as the infimum of the Lagrangian over $\boldsymbol{x}$: $g: \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$,

$\forall\ \tilde{x} \in \mathrm{dom} f$

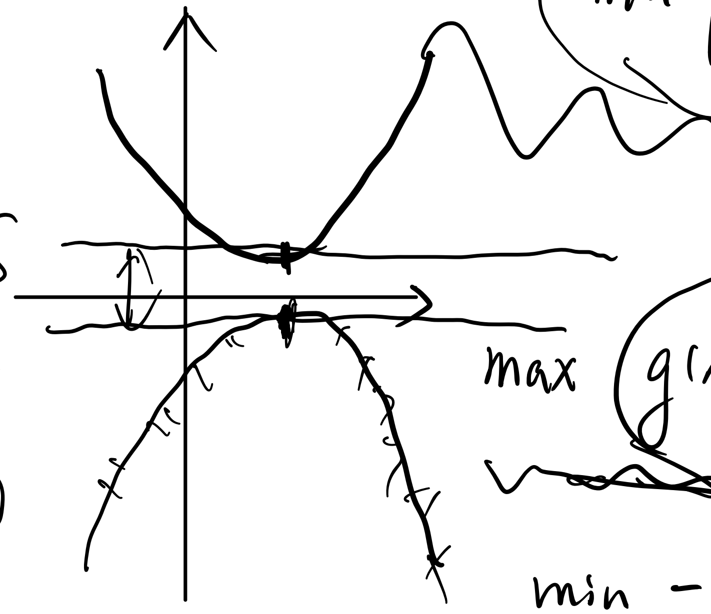
$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$$

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

Δ ≥ 0 ≤ 0 ≤ 0 0 min

- Observe that:
  - the infimum is unconstrained (as opposed to the original constrained minimization problem)
  - g is concave regardless of original problem (infimum of affine functions)
  - g can be $-\infty$ for some $\boldsymbol{\lambda}, \boldsymbol{\nu}$

$\min f(x)$ $\quad \underline{p^*}$



duality gap

$\max g(\lambda, \nu)$ $\quad \underline{d^*}$

$p^* - d^* \geq 0$

$\min -g(\lambda, \nu)$

$\max g(\lambda, \nu) \leq \min f(x)$

Lagrange Dual Function II

- **Lower bound property:** if $\boldsymbol{\lambda} \succeq \mathbf{0}$, then $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^\star$.

Proof.

Suppose $\tilde{\boldsymbol{x}}$ is feasible and $\boldsymbol{\lambda} \succeq \mathbf{0}$. Then,

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

Now choose minimizer of $f_0(\tilde{\boldsymbol{x}})$ over all feasible $\tilde{\boldsymbol{x}}$ to get $p^\star \geq g(\boldsymbol{\lambda}, \boldsymbol{\nu})$. $\square$

- We could try to find the best lower bound by maximizing $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$. This is in fact the dual problem.

Outline

Dual Problem

- The *Lagrange dual problem* is defined as

$$\begin{array}{ll} \underset{\boldsymbol{\lambda},\boldsymbol{\nu}}{\text{maximize}} & g(\boldsymbol{\lambda},\boldsymbol{\nu}) \\ \text{subject to} & \boldsymbol{\lambda} \succeq \mathbf{0} \end{array}$$

- This problem finds the best lower bound on $p^\star$ obtained from the dual function

- It is a convex optimization (maximization of a concave function and linear constraints)

- The optimal value is denoted $d^\star$

- $\boldsymbol{\lambda}, \boldsymbol{\nu}$ are dual feasible if $\boldsymbol{\lambda} \succeq \mathbf{0}$ and $(\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \operatorname{dom} g$ (the latter implicit constraints can be made explicit in problem formulation)

Example: Least-Norm Solution of Linear Equations I

$\langle x, x \rangle$ $\|x\|_F^2$

- Consider the problem

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{x}^T\boldsymbol{x} = \|x\|_2^2 = \mathrm{Tr}(x^T \cdot x) \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

$Ax = b$

- The Lagrangian is

$$L(\boldsymbol{x}, \boldsymbol{\nu}) = \boldsymbol{x}^T\boldsymbol{x} + \boldsymbol{\nu}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$$

- To find the dual function, we need to solve an unconstrained minimization of the Lagrangian. We set the gradient equal to zero

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

$\triangle$

Example: Least-Norm Solution of Linear Equations II

and we plug the solution in L to obtain g:

$$g(\boldsymbol{\nu}) = L(-\frac{1}{2}\boldsymbol{A}^T\boldsymbol{\nu}, \boldsymbol{\nu}) = -\frac{1}{4}\boldsymbol{\nu}^T\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{\nu} - \boldsymbol{b}^T\boldsymbol{\nu}$$

- The function g is, as expected, a concave function of $\boldsymbol{\nu}$.
- From the lower bound property, we have

$$p^\star \geq -\frac{1}{4}\boldsymbol{\nu}^T\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{\nu} - \boldsymbol{b}^T\boldsymbol{\nu} \text{ for all } \boldsymbol{\nu}$$

- The dual problem is the QP

$$\underset{\boldsymbol{\nu}}{\text{maximize}} \quad -\frac{1}{4}\boldsymbol{\nu}^T\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{\nu} - \boldsymbol{b}^T\boldsymbol{\nu}$$

$$-\frac{1}{2} A \cdot A^T \nu - b = 0$$

$$\nu = -(A \cdot A^T)^{-1} \cdot b$$

Example: Standard Form LP I

- Consider the problem

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \quad \boldsymbol{x} \succeq \boldsymbol{0} \end{array}$$

$Gx \leq h$

$Gx + s = h$

- The Lagrangian is

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) &= \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{\nu}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) - \boldsymbol{\lambda}^T\boldsymbol{x} \\ &= (\underbrace{\boldsymbol{c} + \boldsymbol{A}^T\boldsymbol{\nu} - \boldsymbol{\lambda}}_{0})^T\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{\nu} \end{aligned}$$

- L is a linear function of $\boldsymbol{x}$ and it is unbounded if the term multiplying $\boldsymbol{x}$ is nonzero.

$$\begin{bmatrix} A & 0 \\ G & I \end{bmatrix} \begin{bmatrix} x \\ s \end{bmatrix} = \begin{bmatrix} b \\ h \end{bmatrix}$$

Example: Standard Form LP II

$\lambda \geq 0$

- Hence, the dual function is

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

- The function g is a concave function of $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ as it is linear on an affine domain.
- From the lower bound property, we have

$$p^\star \geq -\boldsymbol{b}^T\boldsymbol{\nu} \quad \text{if } \boldsymbol{c} + \boldsymbol{A}^T\boldsymbol{\nu} \succeq \boldsymbol{0}$$

- The dual problem is the LP

$$\begin{array}{ll} \underset{\boldsymbol{\nu}}{\text{maximize}} & -\boldsymbol{b}^T\boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{c} + \boldsymbol{A}^T\boldsymbol{\nu} \succeq \boldsymbol{0} \end{array}$$

Outline

Weak and Strong Duality I

max

- From the lower bound property, we know that $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^\star$ for feasible $(\boldsymbol{\lambda}, \boldsymbol{\nu})$. In particular, for a $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ that solves the dual problem.

- Hence, **weak duality** always holds (even for nonconvex problems):

$$d^\star \leq p^\star$$

- The difference $p^\star - d^\star$ is called **duality gap**.

- Solving the dual problem may be used to find nontrivial lower bounds for difficult problems.

- Even more interesting is when equality is achieved in weak duality. This is called **strong duality**:

$$d^\star = p^\star$$

Weak and Strong Duality II

- Strong duality means that the duality gap is zero.
- Strong duality:
  - is very desirable (we can solve a difficult problem by solving the dual)
  - does not hold in general
  - usually holds for convex problems
  - conditions that guarantee strong duality in convex problems are called **constraint qualifications**.

Slater's Constraint Qualification I

- Slater's constraint qualification is a very simple condition that is satisfied in most cases and ensures strong duality for convex problems.

- Strong duality hold for a convex problem

$$\begin{array}{lll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) & \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0 & i = 1, \cdots, m \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} & \end{array}$$

  if it is strictly feasible, i.e.,

$$\exists\, \boldsymbol{x} \in \operatorname{int} \mathcal{D} : \quad f_i(\boldsymbol{x}) < 0 \quad i = 1, \cdots, m, \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

- There exist many other types of constraint qualifications.

Example: Inequality Form LP

- Consider the problem

$$L(x,\lambda) = c^T x + \lambda^T (Ax - b)$$

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{c}^T \boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b} \end{array}$$

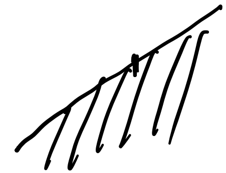
$$(\underbrace{c^T + \lambda^T A}_{0})x - \underline{\lambda^T b}$$

- The dual problem is

$$\begin{array}{ll} \underset{\boldsymbol{\lambda}}{\text{maximize}} & -\boldsymbol{b}^T \boldsymbol{\lambda} \\ \text{subject to} & \boldsymbol{A}^T \boldsymbol{\lambda} + \boldsymbol{c} = \boldsymbol{0}, \quad \boldsymbol{\lambda} \succeq \boldsymbol{0} \end{array}$$

- From Slater's condition: $p^\star = d^\star$ if $\boldsymbol{A}\tilde{\boldsymbol{x}} \prec \boldsymbol{b}$ for some $\tilde{\boldsymbol{x}}$.
- In this case, in fact, $p^\star = d^\star$ except when primal and dual are infeasible.

Example: Convex QP



- Consider the problem (assume $\boldsymbol{P} \succeq \boldsymbol{0}$)

$$
\begin{array}{ll}
\underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x} \quad + \lambda^T (Ax - b) \\
\text{subject to} & \boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b}
\end{array}
$$

$$2Px + A^T \lambda = 0$$

- The dual problem is

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

$$
\begin{array}{ll}
\underset{\boldsymbol{\lambda}}{\text{maximize}} & -\dfrac{1}{4}\boldsymbol{\lambda}^T \boldsymbol{A}\boldsymbol{P}^{-1}\boldsymbol{A}^T\boldsymbol{\lambda} - \boldsymbol{b}^T\boldsymbol{\lambda} \\
\text{subject to} & \boldsymbol{\lambda} \succeq \boldsymbol{0}
\end{array}
$$

- From Slater's condition: $p^\star = d^\star$ if $\boldsymbol{A}\tilde{\boldsymbol{x}} \prec \boldsymbol{b}$ for some $\tilde{\boldsymbol{x}}$.

- In this case, in fact, $p^\star = d^\star$ always.

Complementary Slackness

- Assume strong duality holds, $\boldsymbol{x}^\star$ is primal optimal and $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ is dual optimal. Then

$$
\begin{aligned}
f_0(\boldsymbol{x}^\star) = g(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) &= \inf_{\boldsymbol{x}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^\star f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i^\star h_i(\boldsymbol{x}) \right) \\
&\leq f_0(\boldsymbol{x}^\star) + \underbrace{\sum_{i=1}^{m} \underbrace{\lambda_i^\star}_{\geq 0} \underbrace{\overset{0}{\overset{\|}{f_i}}(\boldsymbol{x}^\star)}_{\leq 0}}_{\leq 0} + \sum_{i=1}^{p} \nu_i^\star \underbrace{h_i(\boldsymbol{x}^\star)}_{\substack{\| \\ 0}} \\
&\leq f_0(\boldsymbol{x}^\star)
\end{aligned}
$$

- Hence, the two inequalities must hold with equality. Implications:
  - $\boldsymbol{x}^\star$ minimizes $L(\boldsymbol{x}, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$
  - $\lambda_i^\star f_i(\boldsymbol{x}^\star) = 0$ for $i = 1, \cdots, m$; this is called **complementary slackness**:

$$
\lambda_i^\star > 0 \Longrightarrow f_i(\boldsymbol{x}^\star) = 0, \quad f_i(\boldsymbol{x}^\star) < 0 \Longrightarrow \lambda_i^\star = 0
$$

Outline

$$\max_{(\lambda,\nu)\in D_g} \inf_{x \in D_f} L(x,\lambda,\nu) =$$

$$\nabla_x L(x,\lambda,\nu) = 0$$

Karush-Kuhn-Tucker (KKT) Conditions

KKT conditions (for differentiable f_i, h_i):

1. primal feasibility:

$$f_i(\boldsymbol{x}) \leq 0, \ i = 1, \cdots, m, \ h_i(\boldsymbol{x}) = 0, \ i = 1, \cdots, p$$

2. dual feasibility: $\boldsymbol{\lambda} \succeq \mathbf{0}$

3. complementary slackness: $\lambda_i^\star f_i(\boldsymbol{x}^\star) = 0$ for $i = 1, \cdots, m$

4. zero gradient of Lagrangian with respect to $\boldsymbol{x}$:

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

KKT condition

- We already known that if strong duality holds and $\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}$ are optimal, then they must satisfy the KKT conditions.
- What about the opposite statement?
- If $\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}$ satisfy the KKT conditions for a convex problem, then they are optimal.

Proof.

From complementary slackness, $f_0(\boldsymbol{x}) = L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ and, from 4th KKT condition and convexity, $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$. Hence, $f_0(\boldsymbol{x}) = g(\boldsymbol{\lambda}, \boldsymbol{\nu})$. □

Theorem

If a problem is convex and Slater's condition is satisfied, then x is optimal if and only if there exists $\boldsymbol{\lambda}, \boldsymbol{\nu}$ that satisfy the KKT conditions.

$$\min_{x \in \mathbb{R}^n} \; -\sum_{i=1}^{n} \log(a_i + x_i), \quad a_i > 0. \qquad \frac{\partial f(x_1, \cdots x_n)}{\partial x}$$

$$\text{s.t.} \quad x \geq 0. \quad \vec{1} \cdot x = 1 \quad \text{simplex}$$

$$L(x, \lambda, \nu) = \underbrace{-\sum_{i=1}^{n} \log(a_i + x_i)} - \lambda^T x + \nu(\vec{1} \cdot x - 1)$$

$$\begin{cases} \dfrac{\partial L}{\partial x} = 0 \\ \lambda \geq 0 \\ x \geq 0, \; \boxed{\vec{1} \cdot x = 1} \\ \lambda^T \cdot x = 0 \end{cases} \Longrightarrow \begin{bmatrix} -\dfrac{1}{a_1 + x_1} \\ -\dfrac{1}{a_2 + x_2} \\ \vdots \\ -\dfrac{1}{a_n + x_n} \end{bmatrix} - \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} + \begin{bmatrix} \nu \\ \nu \\ \vdots \\ \nu \end{bmatrix} = 0$$

$$\Rightarrow \quad -\frac{1}{a_i + x_i} - \lambda_i + \nu = 0$$

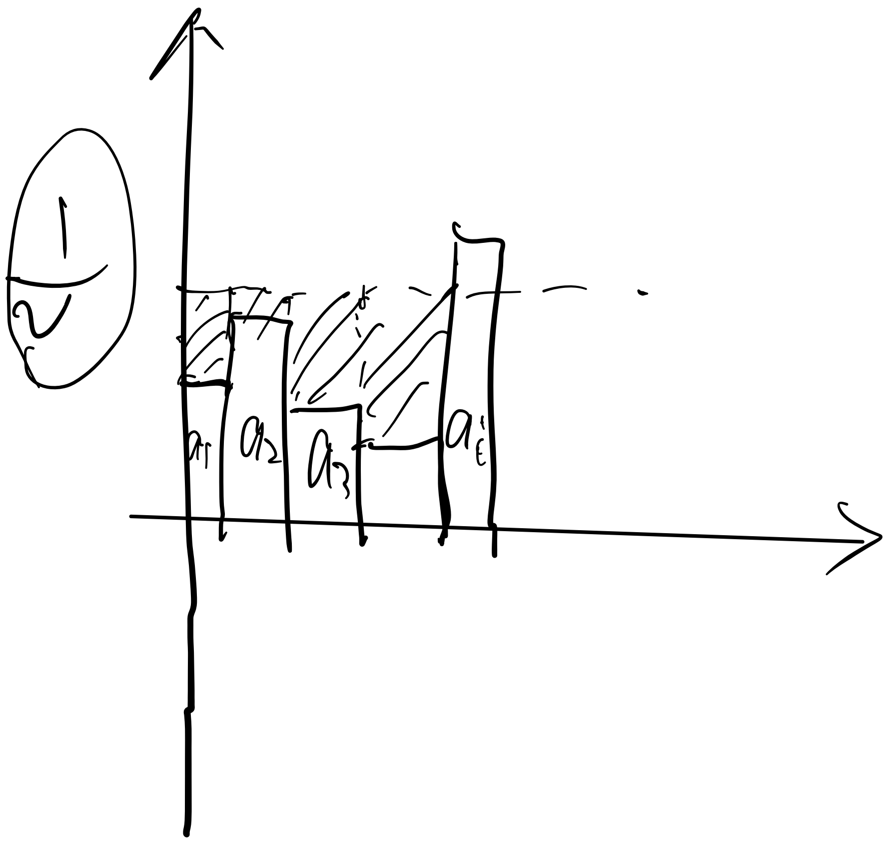
$$\underbrace{\lambda_i = \frac{1}{a_i + x_i} - \nu} \geq 0$$

$$x_i \lambda_i = 0$$

$$\begin{cases} 0 < \nu < \dfrac{1}{a_i} \Rightarrow x_i > 0, \quad \lambda_i = 0. \Rightarrow \underbrace{x_i = \dfrac{1}{\nu} - a_i} \\ \nu = \dfrac{1}{a_i} \Rightarrow x = 0 \\ \nu > \dfrac{1}{a_i} \Rightarrow \lambda_i > 0 \Rightarrow \underbrace{x_i = 0} \end{cases}$$

$$\Rightarrow \quad x_i = \max\left\{0,\ \frac{1}{v} - a_i\right\}$$

$$\sum_{i=1}^{n} x_i = 1 \quad \Rightarrow \quad \sum_{i=1}^{n} \max\left\{0,\ \frac{1}{v} - a_i\right\} = 1$$



Reference

Chapter 5 of:

- Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.