then there is a unique x_y such that $\nabla f_{x_y} = y$, so that

$$f^*(y) = x_y^\top \nabla f_{x_y} - f(x_y),$$

and f^* is differentiable with

$$\nabla f^*_y = x_y.$$

We now return to our optimization problem.

Proposition 50.20. *Consider Problem (P),*

$$\begin{aligned} \text{minimize} \quad & J(v) \\ \text{subject to} \quad & Av \leq b \\ & Cv = d, \end{aligned}$$

with affine inequality and equality constraints (with A an $m \times n$ matrix, C an $p \times n$ matrix, $b \in \mathbb{R}^m$, $d \in \mathbb{R}^p$). The dual function $G(\lambda, \nu)$ is given by

$$G(\lambda, \nu) = \begin{cases} -b^\top \lambda - d^\top \nu - J^*(-A^\top \lambda - C^\top \nu) & \textit{if } -A^\top \lambda - C^\top \nu \in \mathrm{dom}(J^*), \\ -\infty & \textit{otherwise}, \end{cases}$$

for all $\lambda \in \mathbb{R}^m_+$ and all $\nu \in \mathbb{R}^p$, where J^* is the conjugate of J.

Proof. The Lagrangian associated with the above program is

$$\begin{aligned} L(v, \lambda, \nu) &= J(v) + (Av - b)^\top \lambda + (Cv - d)^\top \nu \\ &= -b^\top \lambda - d^\top \nu + J(v) + (A^\top \lambda + C^\top \nu)^\top v, \end{aligned}$$

with $\lambda \in \mathbb{R}^m_+$ and $\nu \in \mathbb{R}^p$. By definition

$$\begin{aligned} G(\lambda, \nu) &= -b^\top \lambda - d^\top \nu + \inf_{v \in \mathbb{R}^n} (J(v) + (A^\top \lambda + C^\top \nu)^\top v) \\ &= -b^\top \lambda - d^\top \nu - \sup_{v \in \mathbb{R}^n} (-(A^\top \lambda + C^\top \nu)^\top v - J(v)) \\ &= -b^\top \lambda - d^\top \nu - J^*(-A^\top \lambda - C^\top \nu). \end{aligned}$$

Therefore, for all $\lambda \in \mathbb{R}^m_+$ and all $\nu \in \mathbb{R}^p$, we have

$$G(\lambda, \nu) = \begin{cases} -b^\top \lambda - d^\top \nu - J^*(-A^\top \lambda - C^\top \nu) & \text{if } -A^\top \lambda - C^\top \nu \in \mathrm{dom}(J^*), \\ -\infty & \text{otherwise}, \end{cases}$$

as claimed. ∎

As application of Proposition 50.20, consider the following example.