and

$$0 \in \partial f(x^*) + \partial g(z^*) + A^\top \lambda^* + B^\top \lambda^*, \qquad (\dagger)$$

Assumption (3) is also equivalent to Conditions (a) and (b) of Theorem 51.41. In particular, our program has an optimal solution (x^*, z^*). By Theorem 51.43, λ^* is maximizer of the dual function $G(\lambda) = \inf_{x,z} L_0(x, z, \lambda)$ and strong duality holds, that is, $G(\lambda^*) = f(x^*) + g(z^*)$ (the duality gap is zero).

We will show after the proof of Theorem 52.1 that Assumption (2) is actually implied by Assumption (3). This allows us to prove a convergence result stronger than the convergence result proven in Boyd et al. [28] (under the exact same assumptions (1) and (3)). In particular, we prove that *all* of the sequences (x^k), (z^k), and (λ^k) converge to optimal solutions $(\widetilde{x}, \widetilde{z})$, and $\widetilde{\lambda}$. The core of our proof is due to Boyd et al. [28], but there are new steps because we have the stronger hypothesis (2).

In Section 52.5, we discuss stopping criteria.

In Section 52.6 we present some applications of ADMM, in particular, minimization of a proper closed convex function f over a closed convex set C in $\mathbb{R}^n$ and quadratic programming. The second example provides one of the best methods for solving quadratic problems, including the SVM problems discussed in Chapter 54, the elastic net method in Section 55.6, and ν-SV regression in Chapter 56.

Section 52.8 gives applications of ADMM to ℓ^1-norm problems, in particular, lasso regularization, which plays an important role in machine learning.

52.1 Dual Ascent

Our goal is to solve the minimization problem, Problem (P),

$$\begin{aligned} &\text{minimize} && J(u) \\ &\text{subject to} && Au = b, \end{aligned}$$

with affine equality constraints (with A an $m \times n$ matrix and $b \in \mathbb{R}^m$). The Lagrangian $L(u, \lambda)$ of Problem (P) is given by

$$L(u, \lambda) = J(u) + \lambda^\top (Au - b).$$

with $\lambda \in \mathbb{R}^m$. From Proposition 50.20, the dual function $G(\lambda) = \inf_{u \in \mathbb{R}^n} L(u, \lambda)$ is given by

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

for all $\lambda \in \mathbb{R}^m$, where J^* is the conjugate of J. Recall that by Definition 50.11, the *conjugate* f^* of a function $f \colon U \to \mathbb{R}$ defined on a subset U of $\mathbb{R}^n$ is the partial function $f^* \colon \mathbb{R}^n \to \mathbb{R}$ defined by

$$f^*(y) = \sup_{x \in U} (y^\top x - f(x)), \quad y \in \mathbb{R}^n.$$