dent. In this case, equations (2.1, 2.2) specialize to

$$\widehat{x} = PC^T \left(CPC^T + R\right)^{-1} y \tag{2.4}$$

$$\widehat{P} = P - PC^T \left(CPC^T + R\right)^{-1} CP \tag{2.5}$$

where $P \triangleq E\left[xx^T\right]$ is the prior covariance of the unobserved variables x. Assuming that P and R are both positive definite and hence invertible, we may use the matrix inversion lemma (see Appendix A.1) to rewrite equations (2.4, 2.5) as

$$\left(P^{-1} + C^T R^{-1} C\right)\widehat{x} = C^T R^{-1} y \tag{2.6}$$

$$\widehat{P} = \left(P^{-1} + C^T R^{-1} C\right)^{-1} \tag{2.7}$$

This information form of the normal equations plays a crucial role in computations involving the structured prior models considered later in this thesis.

Although the normal equations give an explicit formula for determining both $\widehat{x}$ and $\widehat{P}$, directly solving them may require an intractable amount of computation for large–scale estimation problems. Let N_x be the dimension of x, and N_y be the dimension of y. If we assume that R is diagonal but all other matrices are full, then equations (2.4, 2.5) require $\mathcal{O}(N_y^3 + N_x^2 N_y)$ operations, while equations (2.6, 2.7) require $\mathcal{O}(N_x^3)$. For practical problems arising in fields such as image processing and oceanography, $N_x \approx N_y \approx 10^5$ and either formulation is intractable. Also, note that in the absence of special structure, simply storing P or $\widehat{P}$ requires $\mathcal{O}(N_x^2)$ bytes, which is often unreasonably large. These costs motivate the development of structured statistical models which yield efficient estimation procedures.

2.2 Graphical Models

Graphical models provide a powerful general framework for encoding the probabilistic structure of a set of random variables [43, 48]. They are ideally suited to applications where the statistical behavior of a large, complex system may be specified in terms