input: K (covariance matrix), $\mathbf{y}$ (± 1 targets), $p(\mathbf{y}|\mathbf{f})$ (likelihood function)

2: $\mathbf{f} := \mathbf{0}$ initialization

repeat Newton iteration

4: $W := -\nabla\nabla \log p(\mathbf{y}|\mathbf{f})$ eval. W e.g. using eq. (3.15) or (3.16)

$L := \text{cholesky}(I + W^{\frac{1}{2}} K W^{\frac{1}{2}})$ $B = I + W^{\frac{1}{2}} K W^{\frac{1}{2}}$

6: $\mathbf{b} := W\mathbf{f} + \nabla \log p(\mathbf{y}|\mathbf{f})$

$\mathbf{a} := \mathbf{b} - W^{\frac{1}{2}} L^{\top} \backslash (L \backslash (W^{\frac{1}{2}} K \mathbf{b}))$ } eq. (3.18) using eq. (3.27)

8: $\mathbf{f} := K\mathbf{a}$

until convergence objective: $-\frac{1}{2}\mathbf{a}^{\top}\mathbf{f} + \log p(\mathbf{y}|\mathbf{f})$

10: $\log q(\mathbf{y}|X, \theta) := -\frac{1}{2}\mathbf{a}^{\top}\mathbf{f} + \log p(\mathbf{y}|\mathbf{f}) - \sum_i \log L_{ii}$ eq. (3.32)

return: $\hat{\mathbf{f}} := \mathbf{f}$ (post. mode), $\log q(\mathbf{y}|X, \theta)$ (approx. log marg. likelihood)

Algorithm 3.1: Mode-finding for binary Laplace GPC. Commonly used convergence criteria depend on the difference in successive values of the objective function $\Psi(\mathbf{f})$ from eq. (3.12), the magnitude of the gradient vector $\nabla\Psi(\mathbf{f})$ from eq. (3.13) and/or the magnitude of the difference in successive values of $\mathbf{f}$. In a practical implementation one needs to secure against divergence by checking that each iteration leads to an increase in the objective (and trying a smaller step size if not). The computational complexity is dominated by the Cholesky decomposition in line 5 which takes $n^3/6$ operations (times the number of Newton iterations), all other operations are at most quadratic in n.