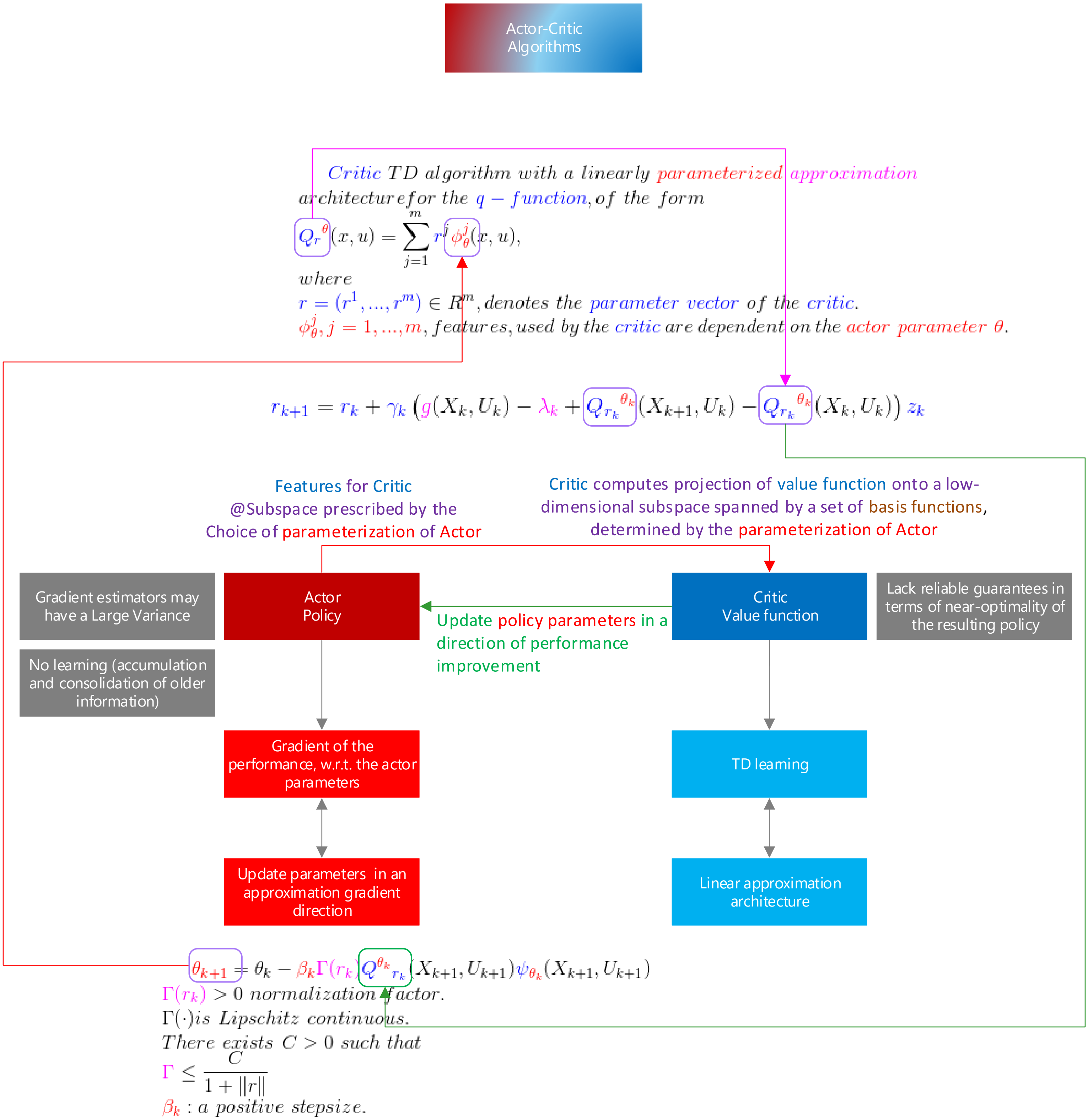
Actor-Critic Algorithms
Critic TD algorithm with a linearly parameterized approximation architecture for the q − function, of the form
$Q_r^{\theta}(x,u) = \sum_{j=1}^{m} r^j \phi_\theta^j(x,u)$,
where
$r = (r^1, ..., r^m) \in R^m$, denotes the parameter vector of the critic.
$\phi_\theta^j, j = 1, ..., m$, features, used by the critic are dependent on the actor parameter θ.
$r_{k+1} = r_k + \gamma_k \left(g(X_k, U_k) - \lambda_k + Q_{r_k}^{\theta_k}(X_{k+1}, U_k) - Q_{r_k}^{\theta_k}(X_k, U_k)\right) z_k$
Features for Critic
@Subspace prescribed by the Choice of parameterization of Actor
Critic computes projection of value function onto a low-dimensional subspace spanned by a set of basis functions, determined by the parameterization of Actor
Gradient estimators may have a Large Variance
No learning (accumulation and consolidation of older information)
Actor Policy
Update policy parameters in a direction of performance improvement
Critic Value function
Lack reliable guarantees in terms of near-optimality of the resulting policy
Gradient of the performance, w.r.t. the actor parameters
TD learning
Update parameters in an approximation gradient direction
Linear approximation architecture
$\theta_{k+1} = \theta_k - \beta_k \Gamma(r_k) Q_{r_k}^{\theta_k}(X_{k+1}, U_{k+1}) \psi_{\theta_k}(X_{k+1}, U_{k+1})$
$\Gamma(r_k) > 0$ normalization factor.
$\Gamma(\cdot)$ is Lipschitz continuous.
There exists $C > 0$ such that
$\Gamma \leq \frac{C}{1 + \|r\|}$
β_k : a positive stepsize.