The 3D space of all Value Functions over 3 states

Path of Dynamic Programming Policy Evaluation

$\boldsymbol{V}^\pi$

$\boldsymbol{w}_{TDE}$

SGD with gradient as
$\nabla_w(\boldsymbol{E}_\pi[(r + \gamma\boldsymbol{V}_w(s') - \boldsymbol{V}_w(s))^2])$

$BE = \boldsymbol{B}^\pi \cdot \boldsymbol{V}_w - \boldsymbol{V}_w$
$PBE = \prod_\phi \cdot \boldsymbol{B}^\pi \cdot \boldsymbol{V}_w - \boldsymbol{V}_w$

MC with linear function approximation gets here (eventually)

$\prod_\phi \cdot \boldsymbol{V}^\pi$
$\boldsymbol{w}_\pi = \arg\min_w d(\boldsymbol{V}^\pi, \boldsymbol{V}_w)$

$\boldsymbol{B}^\pi \cdot \boldsymbol{V}_w$

$\boldsymbol{w}_{PBE} = \arg\min_w d(\prod_\phi \cdot \boldsymbol{B}^\pi \cdot \boldsymbol{V}_w, \boldsymbol{V}_w)$

Bellman Error Vector (BE)

TD fixed-point

Path of sequence of $\prod_\phi \cdot \boldsymbol{B}^\pi$ operations

PBE

$\prod_\phi \cdot \boldsymbol{B}^\pi \cdot \boldsymbol{V}_w$

$\boldsymbol{w}_{BE} = \arg\min_w d(\boldsymbol{B}^\pi \cdot \boldsymbol{V}_w, \boldsymbol{V}_w)$

$\boldsymbol{V}_w$

$\boldsymbol{w}_1$

The subspace of all Value Functions representable as $\boldsymbol{V}_w$

Bellman Error minimizing point
SGD with gradient as
$\nabla_w(\boldsymbol{E}_\pi[r + \gamma\boldsymbol{V}_w(s') - \boldsymbol{V}_w(s)])^2$

$\boldsymbol{w}_2$