BTP - Contextual Linear bandit with linear constraints

(Implementation)

→ $\langle x, y\rangle = x^T y$ $\quad \|x\| = \sqrt{x^T x}$ $\quad$ x = vector (d×1)

x^T → lamda - prob. of selecting each encoder ($x = \pi$) (Try other version too!)

Work: $\epsilon(d \times C)$, $R(d \times 1)$

```
start with uniform x
for 1 to T :
        Sample an encoder using $x_t$
        get 0 or 1 by using that encoder
        assign prob. $y_t$ to channels with 0/1
        $r_t = \langle x_t, R\rangle$ → True
        $c_t = \langle x_t, \epsilon y_t\rangle$ → estimate
        # $\pi_t \cong x_t$ , $\tau$ = tol
        Do Algorithm
```

Algorithm: $x_t (d \times 1)$, r_t, c_t (Scalars)

prereq:

θ^* - R → Normalized s.t $\sum R_i = 1$ (To get s)

$e_0 = x_0 / \|x_0\|$, x_0 is safe action

$[x_{01}\ x_{02} \dots x_{0d}]$ $\quad \sum x_{0i} = 1$

each row of $x_0^T \epsilon < tol$

find x_0

linear prog

(what if no safe action? can we say problem not feasable?)

$S=1$, $L=1$, ($R=1$ (doubtful !!)) → from assumptions

Start with safe policy

$x^T A + \hat{x}^T B + c\, \hat{x}^T \hat{x} < T$

$\hat{x} = x - \underbrace{(x^T e)}_{\text{Scalar}} e$ [e and x are of same dimension]

$x^T R$ — Manimize

$\sum x_i = 1$

Yet to figure out !

$I_{op} = I_{d\times d} - \frac{1}{\|x_0\|^2} x_0 x_0^T$ | $\alpha_c \geq 1$

Sigopt $= \lambda I_{op}$, vopt $= 0$

start:

$C_{opt} = C_t - \frac{\langle x_t, e_0 \rangle}{\|x_0\|} C_0$

$x_{opt} = x_t - \langle x_t, e_0 \rangle e_0$

Sigopt += xopt × $(\text{xopt})^T$

vopt += copt × xopt

Mopt = (Sigopt^{-1}) × vopt

$\tilde{c}_{\pi t} < T$ (below are simplification steps)

$$\Rightarrow \frac{C_0 e_0}{\|x_0\|} z^t + \left(z - e_0 z^T e_0\right)^T M_{opt} + \alpha_c \beta_t \|x_t\|$$

$$\Downarrow$$

$$\left(z^T - e_0^T z e_0^T\right) M_{opt}$$

$$\rightarrow \underbrace{\frac{C_0}{\|x_0\|} z^T e_0}_{e_0^T z} + \underbrace{z^T M_{opt}}_{M_{opt}^T z} + \underbrace{e_0^T z \overset{1\times 1}{(e_0^T M_{opt})}}_{(e_0^T M_{opt}) e_0^T z} + \alpha_c \beta_t \|x_t\|$$

$$\Rightarrow \left(\frac{C_0}{\|x_0\|} e_0^T + M_{opt}^T + e_0^T x_{opt} e_0^T \right) z + \alpha_c \beta_t \|x_{t-1}\|$$

$\le \tau$ → bound

approx

$R^T z$ → maximaize

$\Sigma z_i = 1$ → eq

Notes on Final Implementation:

i) paper assumes

$r_t = \langle x_t, \theta_* \rangle + \varepsilon_t^r$ but in our case wk

$\theta_* =$ Rate & $\varepsilon_t^r = 0$ precisely so avoided the calculations of θ_t

2) To get a safe action the min_tol should be higher than the min_tol for previous algorithms

3) In cases like x_0 or x_t close to $[1,0,0]$ the M_{uopt} is not possible (singular) or blowing up. hence performing pseudo inverse

4) Final eqn. $Az + C \le T$

$Az \le \tau - C$

here $\tau = 0.2 | 0.3$,

$C = \alpha_c \beta_t \|x_{t-1}\| \simeq \alpha_c \cdot 3.2$

and coeff A are mostly +ve or slightly -ve

So for $\alpha_c \ge 1$ as given in paper we are getting an infeasible equation

* For our problem best $\alpha c = 0.001$ but any valule below 0.01 is working

5) The rate is decreasing starting from a high value to adjust error, In prev. algorithms rate used to increase from a lower value may be due to different starting points (no, uniform)