BTP - Contextual Linear bandit with linear constraints (Implementation)

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

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

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

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

$C_t = \langle x_t, E y_t \rangle$ → estimate

$\pi_t \subseteq x_t$, τ = 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

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

each row of $x_0^T E < tol$

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]

$x^T R$ – Maximize

$\sum x_i = 1$

Yet to figure out !

$I_{op} = I_{d \times d} - \frac{1}{\|x_0\|^2} x_0 x_0^T$ | $\alpha_c \geq 1$
$Sig_{opt} = \lambda I_{op}$, $v_{opt} = 0$

<u>start</u>:

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

$x_{opt} = x_t - \langle x_t, e_0 \rangle e_0$
Sig_{opt} += $x_{opt} \times (x_{opt})^T$
v_{opt} += $C_{opt} \times x_{opt}$
$M_{opt} = (Sig_{opt}^{-1}) \times v_{opt}$

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

$$\Rightarrow \frac{C_0 e_0}{\|x_0\|} z^t + \underbrace{(z - e_0 z^T e_0)^T M_{opt}}_{\Downarrow \; (z^T - e_0^T z e_0^T)\, M_{opt}} + \alpha_c \beta_t \|x_t\|$$

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

1×1

$$\Rightarrow \left(\frac{c_0}{\|x_0\|} e_0^T + M_{opt}^T + e_0^T x_{opt} e_0^T \right) z + \alpha_c \beta_t \|\hat{x}_{t-1}\| \leq \tau \rightarrow \text{bound}$$

approx

$R^T z \rightarrow$ maximaize

$\sum z_i = 1 \rightarrow$ eq

Notes on Final Implementation:

1) 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_{vopt} is not possible (singular) or blowing up. hence performing pseudo inverse

4) Final eqn. $Az + c \leq \tau$

$Az \leq \tau - c$

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 $\alpha_c \geq 1$ as given in paper we are getting an infeasible equation

→ For our problem best $\alpha_c = 0.001$ but any value 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 (x_0, uniform)

Development:

$$\hat{x}^T \hat{x}$$

$$\rightarrow \hat{x} = x - \quad e x^T e$$

$$\rightarrow (x - e x^T e)^T (x - e x^T e)$$

$$(x^T - e^T x e^T)(x - e x^T e)$$

$$x^T x - e^T x e^T x - x^T e x^T e + e^T x e^T e x^T e$$

$$\boxed{((e^T e)^2 + 1) x^T x - 2 e^T x e^T x - 2 (e^T x)^2} \times \alpha_c \beta t$$

Full eqn:

$$A x + b x^T x + d e^T x e^T x \leq \tau$$

$$x^T \, b \times I \, x \qquad x^T \boxed{e e^T} x$$

$$\rightarrow \boxed{A x + x^T (b I + d e e^T) x \leq \tau}$$

$f(x): \; A x + x^T C x - \tau \leq 0$	for modelling except!
$f'(x) = A + 2 C x$	
$f''(x) = 2C$ (2×1)	Done!

$$A = \left(\frac{c_0}{\|n_0\|} e_0' + M_{opt}' + e_0' n_{opt} e_0' \right)$$

$$d = -2\alpha c B t \quad , \quad b = \alpha c B t \left(1 + (e^T e)^2\right)$$

$$C = b * I_{(en \times en)} + d * (e e^T)$$

Notes:

→ Most of regret analysis is done on improvement in θ and the the rest assuming $\alpha c \geq 1$

problem!

→ OPB having the same problem

→ Regret Analysis OPB :

$$= \sum_{1}^{T} E_{a \sim \pi^*} [\bar{r}_a] - E_{a \sim \pi t} [\bar{r}_a]$$

$$= \sum_{1}^{T} \sum_{1}^{A} \theta_a \left(\pi_a^* - \pi t_a \right)$$

θ_a → R_a, π_a^* → LP, πt_a → LP

with $E^* P \to U$ contain   with Uca

$\langle U, \pi^* \rangle \leq T$ $\langle Ucat, \pi t \rangle \leq T$

$\langle U, \pi^* \rangle - \langle Ucat, \pi t \rangle \leq T$