A Proof of Fact 2

Proof. Recall that the safety set $S_t(\bar{\delta})$ after iteration t is defined by the following inequalities:

$$S_t(\bar{\delta}) = \Bigg\{ x \in \mathbb{R}^d : \forall i = 1, \dots, m \ \left[[\hat{a}_t^i]^T x - \hat{b}_t^i\right] + \\ + \phi^{-1}(\bar{\delta}/m)\sigma \left\| (\bar{X}_t^T \bar{X}_t)^{-1/2} \begin{bmatrix} x \\ -1 \end{bmatrix} \right\| \leq 0 \Bigg\}. \quad (13)$$

Remember that $\bar{x}_t = \frac{X^T \mathbf{1}}{N}$ is an average of the measured points. Using the inversion formula for block matrix, we obtain

$$(\bar{X}_t^T \bar{X}_t)^{-1} = \begin{bmatrix} X_t^T X_t & -X_t^T \mathbf{1} \\ -\mathbf{1}^T X_t & N_t \end{bmatrix}^{-1} = \begin{bmatrix} R_t & R_t \bar{x}_t \\ \bar{x}_t^T R_t & \frac{1}{N_t} + \bar{x}_t^T R_t \bar{x}_t \end{bmatrix},$$

where

$$R_t = \left[X_t^T X_t - N_t \bar{x}_t \bar{x}_t^T\right]^{-1} = \left[\sum_{j=1}^{N_t} (x_{(j)} - \bar{x}_t)(x_{(j)} - \bar{x}_t)^T\right]^{-1}. \quad (14)$$

Let us denote by $\phi_{\bar{\delta}} = \sigma\phi^{-1}(\bar{\delta}/m)$ and by $\varepsilon_t^i = \hat{b}_t^i - (\hat{a}_t^i)^T x_t$. Then, the i-th inequality in (13) can be rewritten as follows:

$$\sqrt{\frac{\phi_{\bar{\delta}}^2}{N_t} + \phi_{\bar{\delta}}^2 (x - \bar{x}_t)^T R_t (x - \bar{x}_t)} \leq \varepsilon_t^i.$$

Substituting $x = x_t$ to the above and combining the inequalities together, we obtain that the condition $x_t \in S_t(\bar{\delta})$ is equivalent to

$$\phi_{\bar{\delta}} \sqrt{\frac{1}{N_t} + (x_t - \bar{x}_t)^T R_t (x_t - \bar{x}_t)} \leq \min_{i=1,\dots,m} \varepsilon_t^i.$$

∎

B DFS solution proof

Let us recall that for the polytope $D \in \mathbb{R}^d$, by an active set B we denote a set of indices of d linearly independent constraints active in a vertex $V \in \mathbb{R}^d$ of D, i.e., $V = V^B = [A^B]^{-1} b^B$. Here, A^B is a corresponding sub-matrix of A and b^B is the corresponding right-hand-side of the constraint.

Lemma 2. *If $\beta \in \mathcal{E}_t(\bar{\delta})$ then for any vertex V^B and its estimation $\hat{V}_t^B$ defined by the active set B we have that its estimation error is bounded by*

$$\|\hat{V}_t^B - V^B\| \leq \frac{C_{\bar{\delta}}}{\sqrt{N_t}},$$

where $C_{\bar{\delta}} = \frac{2\phi_{\bar{\delta}} d(\Gamma_0 + 1)}{\rho_{\min}[A^B]} \sqrt{\frac{\Gamma_0^2 + 1}{\omega_0^2} + 1}$.

Proof. The vertex estimate $\hat{V}_t^B$ of a polytope is described by the system of linear equations $\hat{A}_t^B x = \hat{b}_t^B$. Since the LSE (Least Squares Estimation) is unbiased,

$$\mathbb{E}\hat{A}_t^B = A^B, \ \mathbb{E}\hat{b}_t^B = b^B.$$

Let us denote by $\zeta_t = \hat{b}_t^B - b^B$ the uncertainty in estimation of b^B, and $G_t = \hat{A}_t^B - A^B$ the uncertainty in estimation of A^B.

Our aim is to bound the error of the vertex estimation $\|\hat{V}_t^B - V^B\|$. Recall that

$$\hat{V}_t^B - V^B = [\hat{A}_t^B]^{-1} \hat{b}_t^B - [A^B]^{-1} b^B = \\ = [A^B + G_t]^{-1}(b^B + \zeta_t) - [A^B]^{-1} b^B.$$

Note that for any matrices A, B it holds that

$$(A + B)^{-1} = A^{-1} - (I + A^{-1}B)^{-1} A^{-1} B A^{-1}.$$

Therefore, we can modify the expression for the $\hat{V}_t^B - V^B$ as follows

$$\hat{V}_t^B - V^B = \\ = \left[[A^B]^{-1} - (I + [A^B]^{-1} G_t)^{-1} [A^B]^{-1} G_t [A^B]^{-1}\right] (b^B + \zeta_t) \\ - [A^B]^{-1} b^B = \\ = [A^B]^{-1} b^B + [A^B]^{-1} \zeta_t - \\ - (I + [A^B]^{-1} G_t)^{-1} [A^B]^{-1} G_t [A^B]^{-1} (b^B + \zeta_t) - \\ - [A^B]^{-1} b^B = \\ = [A^B]^{-1} \zeta_t - (I + [A^B]^{-1} G_t)^{-1} [A^B]^{-1} G_t [A^B]^{-1} (b^B + \zeta_t).$$

The norm of the difference between the vertex V^B of the set D and its estimation can be bounded by

$$\|\hat{V}_t^B - V^B\| \leq \underbrace{\|[A^B]^{-1} \zeta_t\|}_{(a)} + \\ + \underbrace{\|[A^B]^{-1} G_t\|}_{(b)} \underbrace{\|V^B + [A^B]^{-1} \zeta_t\|}_{(c)} \underbrace{\|I + [A^B]^{-1} G_t\|^{-1}}_{(d)}. \quad (15)$$

To obtain the bounds on the terms (a),(b),(c),(d), let us first obtain the bounds on $\|G_t\|$ and $\|\zeta_t\|$.

Assume that for each $i = 1, \dots, m$ $\beta^i \in \mathcal{E}_t^i(\bar{\delta})$, where

$$\mathcal{E}_t^i(\bar{\delta}) = \left\{ z \in R^{d+1} : (\hat{\beta}_t^i - z)^T \Sigma_t^{-1} (\hat{\beta}_t^i - z) \leq \phi^{-1}(\bar{\delta})^2 \right\},$$

i.e., that for any active set B describing the vertex V^B we have $\beta^B \in \mathcal{E}_t^B(\bar{\delta})$. Consequently, $\|\hat{a}_t^i - a^i\|^2 + |\hat{b}^i - b^i|^2 \leq \phi^{-1}(\bar{\delta}) \|\Sigma_t^{1/2}\|$. Then, for each row of G_t we have

$\|\hat{a}_t^i - a^i\| \leq \phi^{-1}(\bar{\delta})\|\Sigma_t^{1/2}\|$, and for each element of ζ_t we have $|\hat{b}_t^i - b^i| \leq \phi^{-1}(\bar{\delta})\|\Sigma_t^{1/2}\|$. Hence, for $\|G_t\|$ we obtain

$$\|G_t\| \leq \|G\|_F = \sqrt{\sum_{i\in B}\|\hat{a}_t^i - a^i\|_2^2} \leq \sqrt{d}\phi^{-1}(\bar{\delta})\|\Sigma_t^{1/2}\|. \tag{16}$$

Similarly, we obtain a bound on $\|\zeta_t\|$:

$$\|\zeta_t\| = \sqrt{\sum_{i\in B}(\hat{b}_t^i - b^i)^2} \leq \sqrt{d}\phi^{-1}(\bar{\delta})\|\Sigma_t^{1/2}\|. \tag{17}$$

For the LSE covariance matrix norm $\|\Sigma_t^{1/2}\|$ we have

$$\|\Sigma_t^{1/2}\| = \sigma\|(\bar{X}_t^T\bar{X}_t)^{-1}\|^{1/2} =$$
$$= \sigma\left\|\begin{bmatrix} I \\ \bar{x}_t^T\end{bmatrix} R_t \begin{bmatrix} I & \bar{x}_t\end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1/N_t\end{bmatrix}\right\|^{1/2} \leq$$
$$\leq \sigma\sqrt{\|R_t\|\|\bar{x}_t\|^2 + \|R_t\| + \frac{1}{N_t}},$$

where R_t was defined in (14).

Note that we make measurements as it is described in Step 4 of the SFW algorithm, i.e., we make measurements at all coordinate directions within small step size ω_0 from points generated by the method. Then, each new $2d$ measurements result in addition of a matrix $\sum_{j=N_t+1}^{N_t+2d}(x_{(j)} - \bar{x})(x_{(j)} - \bar{x})^T \succeq \omega_0^2 I$ to the matrix $R_t^{-1} = \sum_{j=1}^{N_t}(x_{(j)} - \bar{x}_t)(x_{(j)} - \bar{x}_t)^T$. Hence, $R_t^{-1} \succeq \frac{N_t\omega_0^2}{2d}I$. Hence, the minimal eigenvalue of the covariance matrix R_t^{-1} is bounded from below by the value $\lambda_{\min}(R_t^{-1}) \geq \frac{N_t\omega_0^2}{d}$. Thus, we obtain the following bound on the norm of R_t:

$$\|R_t\| \leq \frac{d}{N_t\omega_0^2}. \tag{18}$$

Recall that $\|\bar{x}_t\| \leq \Gamma_0$. It follows that

$$\|\Sigma_t^{1/2}\| \leq \sigma\sqrt{\|R_t\|\|\bar{x}_t\|^2 + \|R_t\| + \frac{1}{N_t}} \leq$$
$$\leq \sigma\sqrt{\frac{d}{N_t\omega_0^2}\|\bar{x}_t\|^2 + \frac{d}{N_t\omega_0^2} + \frac{1}{N_t}} \leq$$
$$\leq \frac{\sigma\sqrt{d}\sqrt{\frac{\Gamma_0^2+1}{\omega_0^2} + \frac{1}{d}}}{\sqrt{N_t}} \leq \frac{\sigma\sqrt{d}\sqrt{\frac{\Gamma_0^2+1}{\omega_0^2} + 1}}{\sqrt{N_t}}. \tag{19}$$

In order to bound terms (a),(b),(c),(d) in inequality (15), let us also bound the norm of the matrix $[A^B]^{-1}$:

$$\|[A^B]^{-1}\| = \rho_{\max}([A^B]^{-1}) = \frac{1}{\rho_{\min}[A^B]} \leq \frac{1}{\rho_{\min}(D)}. \tag{20}$$

Then, combining inequalities (16),(17),(19),(20), we bound terms (a) and (b) as follows:

$$(a): \left\|[A^B]^{-1}\zeta_t\right\| \leq \|[A^B]^{-1}\|\|\zeta_t\| \leq \frac{U}{\sqrt{N_t}},$$
$$(b): \left\|[A^B]^{-1}G_t\right\| \leq \|[A^B]^{-1}\|\|G_t\| \leq \frac{U}{\sqrt{N_t}},$$

where U we defined by

$$U = \frac{\phi_{\bar{\delta}} d}{\rho_{\min}(D)}\sqrt{\frac{\Gamma_0^2+1}{\omega_0^2} + 1}.$$

Further, let us bound term (d). For $N_t \geq 4U^2$ it holds that

$$\left\|[A^B]^{-1}G_t\right\| \leq \frac{U}{\sqrt{N_t}} \leq \frac{1}{2},$$
$$\left\|[A^B]^{-1}\zeta_t\right\| \leq \frac{U}{\sqrt{N_t}} \leq \frac{1}{2}.$$

As such, for $N_t \geq 4U^2$ we have

$$\left\|I + [A^B]^{-1}G_t\right\|^{-1} \leq 2. \tag{21}$$

Finally, term (c) we can bound as follows:

$$\left\|V^B + [A^B]^{-1}\zeta_t\right\| \leq \Gamma_0 + \frac{U}{\sqrt{N_t}}. \tag{22}$$

Combining these all together, we obtain

$$\|\hat{V}_t^B - V^B\| \leq \frac{U}{\sqrt{N_t}} + 2\frac{U}{\sqrt{N_t}}\left(\Gamma_0 + \frac{U}{\sqrt{N_t}}\right) \leq$$
$$\leq 2\frac{U}{\sqrt{N_t}}(\Gamma_0 + 1) = \frac{C_{\bar{\delta}}}{\sqrt{N_t}},$$

where

$$C_{\bar{\delta}} = 2U(\Gamma_0 + 1) = \frac{2\phi_{\bar{\delta}} d(\Gamma_0 + 1)}{\rho_{\min}(D)}\sqrt{\frac{\Gamma_0^2+1}{\omega_0^2} + 1}.$$

Since $N_t = C_n t^2(\ln t^2) \geq C_n$, above bound holds under the proper choice of the constant

$$C_n \geq 4U^2 = \frac{C_{\bar{\delta}}^2}{(\Gamma_0 + 1)^2} = \frac{4d^2\phi_{\bar{\delta}}^2}{\rho_{\min}^2(D)}\left(\frac{\Gamma_0^2+1}{\omega_0^2} + 1\right).$$

■

Proposition 1

Proof. Let us find how far is the solution of the random linear DFS from its expected value. Estimated DFS is a linear program defined by

$$\hat{s}_t = \arg\min_{\hat{A}_t x \leq \hat{b}_t} \langle c, x\rangle,$$

where $c = \nabla f(x_t)$. Any solution of such a linear program is a vertex (or convex hull of vertices) of the polytope $\hat{A}_t x \leq \hat{b}_t$. Let $B_2 = \{i_1, \dots, i_d\}$ be a subset of d indices corresponding to linear independent active constraints (which form a basis) at the estimated DFS solution point. And correspondingly, let $B_1 = \{j_1, \dots, j_d\}$ be a subset of indices active at the true DFS solution point.

Assume that V^1 is a true solution of DFS

$$V^1 = \arg\min_{s \in D} \langle c, s \rangle = (A^{B_1})^{-1} b^{B_1}$$

(see Lemma 1 for the definition), and correspondingly

$$\hat{V}^1 = (\hat{A}_t^{B_1})^{-1} \hat{b}_t^{B_1}.$$

Assume also that $\hat{V}_t^2$ is the estimated solution of DFS

$$\hat{V}_t^2 = \arg\min_{s \in \hat{D}_t} \langle c, s \rangle = (\hat{A}_t^{B_2})^{-1} \hat{b}_t^{B_2},$$

and correspondingly $V^2 = (A^{B_2})^{-1} b^{B_2}$. We have that $\hat{V}_t^1 \in \hat{D}_t, V^2 \in D$ is always the case starting from some N_t. Then, from the definitions above it follows that

$$c^T \hat{V}_t^2 \leq c^T \hat{V}_t^1,$$
$$c^T V^2 \leq c^T V^1.$$

From Lemma 2 for the vertices V^1, V^2 corresponding to the active sets B_1, B_2 we have that if $\beta \in \mathcal{E}_t(\bar{\delta}) \cap \mathcal{E}_{t+1}(\bar{\delta})$, then $\|\hat{V}_t^1 - V^1\| \leq \frac{C_{\bar{\delta}}}{\sqrt{N_t}}$ and $\|\hat{V}_t^2 - V^2\| \leq \frac{C_{\bar{\delta}}}{\sqrt{N_t}}$. Hence, we have

$$c^T \hat{V}_t^2 - \|c\| \frac{C_{\bar{\delta}}}{\sqrt{N_t}} \leq c^T \hat{V}_t^1 - \|c\| \frac{C_{\bar{\delta}}}{\sqrt{N_t}}$$
$$\leq c^T V^1 \leq c^T V^2 \leq c^T \hat{V}_t^2 + \|c\| \frac{C_{\bar{\delta}}}{\sqrt{N_t}}.$$

Thus, we obtain that if $\beta \in \mathcal{E}_t(\bar{\delta}) \cap \mathcal{E}_{t+1}(\bar{\delta})$, then $E_t = c^T(\hat{s}_t - s_t) = \|c^T \hat{V}_t^2 - c^T V^1\| \leq \|c\| \frac{C_{\bar{\delta}}}{\sqrt{N_t}}$.

Note that $\|c\| \leq M$, where M is the Lipschitz constant of the objective. Thus, if $\beta \in \mathcal{E}_t(\bar{\delta}) \cap \mathcal{E}_{t+1}(\bar{\delta})$, then $E_t \leq \frac{C_{\bar{\delta}} M}{\sqrt{N_t}}$, i.e.,

$$\mathbb{P}\left\{ E_t \leq \frac{C_{\bar{\delta}} M}{\sqrt{N_t}} \right\} \geq 1 - 2\bar{\delta}.$$

■

C Proof of Lemma 1

First, we provide some preliminary lemmas for the proof of Lemma 1.

Let us denote by $\breve{x}_t = x_t - \bar{x}_t$, $\Delta_t^k = \hat{s}_t - x_k$, and recall that $\varepsilon_t^i = \hat{b}_t^i - [\hat{a}_t^i]^T x_t$. Let us fix some vertex V of the polytope D corresponding to the basic active set B. By $\hat{V}_t$ and $\hat{V}_{t-1}$ we call the estimates of V based on the parameter estimations $\hat{\beta}_t, \hat{\beta}_{t-1}$.

Lemma 3. *If $\beta \in \mathcal{E}_t(\bar{\delta}) \cap \mathcal{E}_{t-1}(\bar{\delta})$ holds, then we have*

$$\min_i \langle \hat{a}_t^i, \Delta_t^t \rangle \geq (1 - \gamma_{t-1}) \min_i \langle \hat{a}_t^i, \Delta_t^{t-1} \rangle - $$
$$- \frac{2\gamma_{t-1} \max_i \|\hat{a}_t^i\| C_{\bar{\delta}}}{\sqrt{N_{t-1}}}.$$

Proof.

$$\forall \hat{s}_t : \min_i \langle \hat{a}_t^i, \Delta_t^t \rangle = \min_i \langle \hat{a}_t^i, \hat{s}_t - x_t \rangle$$
$$= \min_{i \in B_t} \langle \hat{a}_t^i, \hat{s}_t - x_{t-1} - \gamma_{t-1}(\hat{s}_{t-1} - x_{t-1}) \rangle$$
$$= \min_{i \in B_t} \langle \hat{a}_t^i, (1 - \gamma_{t-1})(\hat{s}_t - x_{t-1}) +$$
$$+ \gamma_{t-1}(\hat{s}_{t-1} - \hat{s}_t) \rangle =$$

(Let $j = \arg\min_{i \in B_t} \langle \hat{a}_t^i, \Delta_t^{t-1} \rangle$)

$$= (1 - \gamma_{t-1}) \langle \hat{a}_t^j, \hat{s}_t - x_{t-1} \rangle +$$
$$+ \gamma_{t-1} \langle \hat{a}_t^j, \hat{s}_{t-1} - \hat{s}_t \rangle.$$

Let us fix V as a vertex of D corresponding to active set B_{t-1}, then $\hat{s}_{t-1} = \hat{V}_{t-1}$ and $\langle \hat{a}_t^j, \hat{s}_t \rangle = 0 \leq \langle \hat{a}_t^j, \hat{V}_t \rangle$, thus we have

$$\min_i \langle \hat{a}_t^i, \Delta_t^t \rangle \geq$$
$$\geq (1 - \gamma_{t-1}) \langle \hat{a}_t^j, \hat{s}_t - x_{t-1} \rangle + \gamma_{t-1} \langle \hat{a}_t^j, \hat{V}_{t-1} - \hat{V}_t \rangle$$
$$\geq (1 - \gamma_{t-1}) \min_i \langle \hat{a}_t^i, \hat{s}_t - x_{t-1} \rangle -$$
$$\gamma_{t-1} \max_i \|\hat{a}_t^i\| \|\hat{V}_t - \hat{V}_{t-1}\| \quad (23)$$

Using the result of Proposition 1 we can obtain

$$\|\hat{V}_{t+1} - \hat{V}_t\| \leq \|\hat{V}_{t+1} - V\| + \|V - \hat{V}_t\|$$
$$\leq \|\hat{V}_{t+1} - V\| + \|V - \hat{V}_t\| \leq$$
$$\leq 2\|\hat{V}_t - V\| \leq \frac{C_{\bar{\delta}}}{\sqrt{N_t}} + \frac{C_{\bar{\delta}}}{\sqrt{N_{t+1}}}. \quad (24)$$

Then, combining (24) with (23), we have

$$\min_i \langle \hat{a}_t^i, \Delta_t^t \rangle \geq$$
$$\geq (1 - \gamma_{t-1}) \min_i \langle \hat{a}_t^i, \Delta_t^{t-1} \rangle - \frac{2\gamma_{t-1} \max_i \|\hat{a}_t^i\| C_{\bar{\delta}}}{\sqrt{N_{t-1}}}.$$

■

Lemma 3 above is an induction step in the proof of Lemma 4. Lemma 4 below bounds the fastest rate of decreasing the distance to the boundaries of D for the SCFW algorithm. Recall that $\mathcal{F}_t = \{\beta \in \cap_{k=0}^t \mathcal{E}_k(\bar{\delta})\}$.

Lemma 4. *If $\mathcal{F}_t$ holds, then we have*

$$\min_i \langle \hat{a}_t^i, \Delta_t^t \rangle \geq \frac{\min_i \langle \hat{a}_t^i, \Delta_t^0 \rangle}{t+2} \left(1 - \frac{C_{\bar{\delta}} \ln \ln t \max_i \|\hat{a}_t^i\|}{\sqrt{C_n} \min_i \langle \hat{a}_t^i, \bar{\Delta}_0 \rangle} \right).$$

Proof. By induction, from Lemma 3 we have

$$\min_i \langle \hat{a}_t^i, \hat{s}_t - x_t \rangle \geq \prod_{j=0}^{t-1}(1-\gamma_j)\min_i \langle \hat{a}_t^i, x_0 - \hat{s}_t\rangle - \\ - \sum_{j=0}^{t-1} \frac{2C_{\bar{\delta}}\gamma_j}{\sqrt{N_j}} \max_i \|\hat{a}_t^i\| \prod_{k=j}^{t-1}(1-\gamma_k).$$

Note that $1-\gamma_k = \frac{k+1}{k+2}$, and

$$\prod_{k=j}^{t-1}(1-\gamma_k) = \frac{(t)!/(j+1)!}{(t+1)!/(j+2)!} = \frac{j+2}{t+1}.$$

Thus, $\min_i \langle \hat{a}_t^i, \hat{s}_t - x_t \rangle \geq$

$$\geq \frac{1}{t+1}\min_i \langle \hat{a}_t^i, x_0 - \hat{s}_t\rangle - \sum_{j=0}^{t-1} \frac{j+2}{t+1}\frac{C_{\bar{\delta}}\gamma_j}{\sqrt{N_j}} \max_i \|\hat{a}_t^i\| =$$

$$= \frac{1}{t+1}\min_i \langle \hat{a}_t^i, x_0 - \hat{s}_t\rangle - \frac{1}{t+1}\sum_{j=0}^{t-1} \frac{C_{\bar{\delta}}}{\sqrt{N_j}} \max_i \|\hat{a}_t^i\|.$$

Recall that $N_t = C_n t^2 (\ln t)^2$, hence we have

$$\varepsilon_t^i \geq \min_j \langle \hat{a}_t^j, \hat{s}_t - x_t\rangle \geq \frac{1}{t+2}\min_j \langle \hat{a}_t^j, \hat{s}_t - x_0\rangle - \\ - \frac{1}{t+2}\sum_{j=0}^{t} \frac{\sqrt{2\ln(j+1) + \ln 1/\bar{\delta}}\,C_{\bar{\delta}}}{\sqrt{C_n}(j+1)\ln(j+1)} \max_j \|\hat{a}_t^j\| = \\ = \frac{1}{t+2}\left(\min_j \langle \hat{a}_t^j, \Delta_t^0\rangle - \frac{C_{\bar{\delta}}\ln(\ln t)}{\sqrt{C_n}}\max_j \|\hat{a}_t^j\|\right),$$

where $\Delta_t^0 = \hat{s}_t - x_0$.

■

With Lemmas 3 and 4 in place, we are ready to prove Lemma 1.

C.1 Lemma 1

If $\beta \in \mathcal{E}_k(\bar{\delta})$ for $k = 1,\ldots,m$ and $n_t = 4C_n t(\ln t)^2$, with the constant parameter C_n satisfying

$$C_n \geq C_{\bar{\delta}}^2 \max\left\{\frac{4(\ln\ln T)^2 L_A^2}{[\varepsilon_0]^2}, \frac{1}{(\Gamma_0+1)^2}\right\},$$

$$\text{where } C_{\bar{\delta}} = \frac{2\phi_{\bar{\delta}} d(\Gamma_0+1)}{\rho_{\min}(D)}\sqrt{\frac{\Gamma_0^2+1}{\omega_0^2}+1},$$

then $x_t \in S_t(\bar{\delta})$. Furthermore, the total number of measurements then satisfies $N_t = C_n t^2(\ln t)^2$.

Proof. From Fact 2, the condition $x_t \in S_t(\bar{\delta})$ is equal to

$$\frac{\phi_{\bar{\delta}}^2}{N_t} + \phi_{\bar{\delta}}^2 (x_t - \bar{x}_t)^T R_t (x_t - \bar{x}_t) \leq \min_i [\varepsilon_t^i]^2.$$

From the bound on $\|R_t\|$ given in (18) and knowing that Γ is a diameter of the set, we have

$$\frac{\phi_{\bar{\delta}}^2}{N_t} + \phi_{\bar{\delta}}^2 (x_t - \bar{x}_t)^T R_t (x_t - \bar{x}_t) \leq \frac{\phi_{\bar{\delta}}^2\left(1+\frac{d\Gamma^2}{\omega_0^2}\right)}{N_t}.$$

From Lemma 4 we have

$$[\varepsilon_t^i]^2 \geq \frac{1}{(t+2)^2}\left(\min_i \langle \hat{a}_t^i, \Delta_t^0\rangle - \frac{C_{\bar{\delta}}\ln(\ln t)}{\sqrt{C_n}}\max_i \|\hat{a}_t^i\|\right)^2. \tag{25}$$

Hence, we can guarantee that $x_t \in S_t(\bar{\delta})$ if

$$N_t \geq \frac{(t+2)^2\phi_{\bar{\delta}}^2\left(1+\frac{d\Gamma^2}{\omega_0^2}\right)}{\left(\min_i \langle \hat{a}_t^i, \Delta_t^0\rangle - \frac{C_{\bar{\delta}}\ln(\ln t)}{\sqrt{C_n}}\max_i \|\hat{a}_t^i\|\right)^2}. \tag{26}$$

We denote by $L_A = \max_i \|a_i\|$. Let us derive how far are $\min_i \langle \hat{a}_t^i, \Delta_t^0\rangle$ from ε_0 and $\max_i \|\hat{a}_t^i\|$ from L_A. These are needed for obtaining a bound on the denominator above. If $C_n \geq \frac{C_{\bar{\delta}}^2}{(\Gamma_0+1)^2}$, then with probability $\mathbb{P} \geq 1-\bar{\delta}$ we have $\|\Delta_t^0 - \Delta_0\| \leq \frac{C_{\bar{\delta}}}{\sqrt{N_t}}$. We also can bound the difference $\|\hat{a}_t^i - a^i\|$ by $\|\hat{a}_t^i - a^i\| \leq \phi^{-1}(\bar{\delta})\|\Sigma^{1/2}\| \leq \frac{C_{\bar{\delta}}}{\sqrt{N_t}}\frac{1}{\sqrt{d}\rho_{\min}(D)(\Gamma_0+1)}$. The second inequality follows from (19) and definition of $C_{\bar{\delta}}$ (11).

Combining above inequalities together with the bound (26) on N_t we can conclude the following. If

$$C_n \geq \frac{4C_{\bar{\delta}}^2(\ln\ln T)^2 L_A^2}{\min_i [\varepsilon_0^i]^2},$$

then we can guarantee that $x_t \in S_t(\bar{\delta})$ by requiring

$$N_t \geq \frac{(t+2)^2\phi_{\bar{\delta}}^2\left(1+\frac{d\Gamma^2}{\omega_0^2}\right)}{\min_i [\varepsilon_0^i]^2}.$$

Since $n_t = C_n(t+1)(\ln(t+2))^2$ and $N_t = \sum_{k=0}^t n_k$, we obtain that

$$N_t \geq C_n (t+1)^2 (\ln(t+2))^2.$$

Hence, $C_n \geq \frac{\phi_{\bar{\delta}}^2\left(1+\frac{d\Gamma_0^2}{\omega_0^2}\right)}{(\ln(t+2))^2\min_i[\varepsilon_0^i]^2}$ is enough to ensure that $x_t \in S_t(\bar{\delta})$. Note that $C_{\bar{\delta}} \geq \phi_{\bar{\delta}}^2\left(1+\frac{d\Gamma^2}{\omega_0^2}\right)$. Thus, under the proper choice of constant parameter C_n:

$$C_n \geq \max\left\{\frac{4C_{\bar{\delta}}^2(\ln\ln T)^2 L_A^2}{[\varepsilon_0]^2}, \frac{C_{\bar{\delta}}^2}{(\Gamma_0+1)^2}\right\}$$

we conclude that $x_t \in S_t(\bar{\delta})$. ■

Remark Note that if we use a step size as in classical FW $\gamma_t = \frac{2}{t+2}$ or in more general form $\gamma_t = \frac{l}{t+l}$ then we obtain that the distance to the boundaries $\min_i[\varepsilon_t^i]$ will decrease with rate at most $\prod_{k=0}^{t}(1-\gamma_k) = \frac{l!}{t\cdot\ldots\cdot(t+l)} = O(\frac{1}{t^l})$ instead of (25) and will reach this bound e.g. in the case if the algorithm always moves in the same direction towards the boundary. This implies that in order to keep the convergence rate as in original FW at the same time satisfying $x_t \in S_t(\bar{\delta})$, due to Fact 2 we have to reduce the uncertainty of the boundaries faster, i.e., we need to take more measurements at each iteration.

D Theorem 2

Proof. Let μ_t denote a constant such that $E_t(\delta) = \frac{1}{2}\mu_t\gamma_t C_f$. Then we have

$$\langle \hat{s}, \nabla f(x_t)\rangle \leq \min_{s\in D}\langle s, \nabla f(x_t)\rangle + \frac{1}{2}\mu_t\gamma_t C_f.$$

For the proof we refer to the following result from Jaggi (2013). This result holds in our setting since we use the same notions of g_t and s_t defined in (7).

Lemma 5. *(Lemma 5 Jaggi (2013)) For a step $x_{t+1} = x_t + \gamma(\hat{s} - x_t)$ with the arbitrary step-size $\gamma \in [0,1]$, it holds that*

$$f(x_{t+1}) \leq f(x_t) - \gamma g_t + \frac{\gamma^2}{2}C_f(1+\mu_t),$$

if $\hat{s}$ is an approximate linear minimizer, i.e.

$$\langle \hat{s}, \nabla f(x_t)\rangle \leq \min_{\bar{s}\in D}\langle \bar{s}, \nabla f(x_t)\rangle + \frac{1}{2}\mu_t\gamma C_f.$$

The step-size of the SCFW algorithm is equal to $\gamma_t = \frac{1}{t+2}$. Let us define h_t as follows

$$h_t = h(x_t) = f(x_t) - f(x_*).$$

Then we obtain that

$$\begin{aligned} h_{t+1} &\leq h_t - \gamma_t g_t + \gamma_t^2\frac{C_f}{2}(1+\mu_t) \\ &\leq h_t - \gamma_t h_t + \gamma_t^2\frac{C_f}{2}(1+\mu_t) \\ &= (1-\gamma_t)h_t + \gamma_t^2\frac{C_f}{2}(1+\mu_t). \end{aligned}$$

If we continue in the same manner, we obtain

$$\begin{aligned} h_{t+1} &\leq \prod_{i=0}^{t}(1-\gamma_i)h_0 + \sum_{k=0}^{t}\gamma_k^2\frac{C_f}{2}(1+\mu_k)\prod_{i=k}^{t}(1-\gamma_i) \\ &= \prod_{i=0}^{t}\frac{i+1}{i+2}h_0 + \sum_{k=0}^{t}\frac{1}{(k+2)^2}\frac{C_f}{2}(1+\mu_k)\prod_{i=k}^{t}\frac{i}{i+2} \\ &= \frac{1}{t+2}h_0 + \sum_{k=0}^{t}\frac{1}{(k+2)^2}\frac{C_f(1+\mu_k)}{2}\frac{(t+1)!(k+2)!}{(t+2)!(k+1)!} \\ &= \frac{1}{t+2}\left(h_0 + \sum_{k=0}^{t}\frac{1}{(k+2)}\frac{C_f(1+\mu_k)}{2}\right). \end{aligned}$$

Recall that $E_t(\bar{\delta})$ denotes the upper bound on E_t with the confidence level $1-\bar{\delta}$.

Due to Proposition 1, we have

$$E_k(\bar{\delta}) = \frac{MC_{\bar{\delta}}}{\sqrt{N_k}} = \frac{MC_{\bar{\delta}}}{\sqrt{C_n}(k+2)\ln(k+2)}.$$

Hence, we obtain that

$$\mu_k = \frac{2E_k(\bar{\delta})(k+2)}{C_f} = \frac{2MC_{\bar{\delta}}}{C_f\sqrt{C_n}\ln(k+2)}.$$

Therefore, we obtain

$$\begin{aligned} h_{t+1} &\leq \frac{h_0 + \ln(t+2)\frac{C_f}{2} + \sum_{k=0}^{t}\frac{C_f\mu_k}{2}}{t+2} = \\ &= \frac{h_0 + \ln(t+2)\frac{C_f}{2} + \ln\ln(t+2)\frac{C'}{2}}{t+2}, \end{aligned}$$

where $C' = \frac{MC_{\bar{\delta}}}{\sqrt{C_n}}$. ■

Corollary 1

Proof. Note that we can bound $\phi^{-1}(\bar{\delta})^2$ (Laurent and Massart, 2000) as follows

$$\phi^{-1}(\bar{\delta})^2 \leq d+1+2\log\left(\frac{1}{\bar{\delta}}\right) + 2\sqrt{(d+1)\log\left(\frac{1}{\bar{\delta}}\right)}.$$

Hence, we have

$$\begin{aligned} \phi_{\bar{\delta}} &\leq \sigma\sqrt{d+1+2\log\left(\frac{1}{\bar{\delta}}\right) + 2\sqrt{(d+1)\log\left(\frac{1}{\bar{\delta}}\right)}} \\ &\leq \sigma\left(\sqrt{d+1} + 2\sqrt{\log\left(\frac{1}{\bar{\delta}}\right)}\right) \\ &\leq \sigma\max\left\{\sqrt{d+1}, 2\sqrt{\log\left(\frac{1}{\bar{\delta}}\right)}\right\} = O\left(\sqrt{\ln\frac{1}{\bar{\delta}}}\right). \end{aligned}$$

■