Supplementary for

Provably Optimal Algorithms for Generalized Linear Contextual Bandits

A. Proof of Theorem 1

In the following, for simplicity, we will drop the subscript n when there is no ambiguity. Therefore, V_n is denoted V and so on.

To prove normality-type results of the maximum likelihood estimator $\hat{\theta}$, typically we first show the $n^{-1/2}$-consistency of $\hat{\theta}$ to θ^*. Then, by using a second-order Taylor expansion or Newton-step, we can prove the desired normality of $\hat{\theta}$. More details can be found in standard textbooks such as Van der Vaart (2000).

Since m is twice differentiable with $\ddot{m} \geq 0$, the maximum-likelihood estimation can be written as the solution to the following equation

$$\sum_{i=1}^{n} \left(Y_i - \mu(X_i'\theta)\right) X_i = 0\,. \tag{15}$$

Define $G(\theta) := \sum_{i=1}^{n} \left(\mu(X_i'\theta) - \mu(X_i'\theta^*)\right) X_i$, and we have

$$G(\theta^*) = 0 \;\text{ and }\; G(\hat{\theta}) = \sum_{i=1}^{n} \epsilon_i X_i\,, \tag{16}$$

where the noise ϵ_i is defined in (1). For convenience, define $Z := G(\hat{\theta}) = \sum_{i=1}^{n} \epsilon_i X_i$.

Step 1: Consistency of $\hat{\theta}$. We first prove the consistency of $\hat{\theta}$. For any $\theta_1, \theta_2 \in \mathbb{R}^d$, mean value theorem implies that there exists some $\bar{\theta} = v\theta_1 + (1-v)\theta_2$ with $0 < v < 1$, such that

$$G(\theta_1) - G(\theta_2) = \left[\sum_{i=1}^{n} \dot{\mu}(X_i'\bar{\theta}) X_i X_i'\right](\theta_1 - \theta_2) := F(\bar{\theta})(\theta_1 - \theta_2) \tag{17}$$

Since $\dot{\mu} > 0$ and $\lambda_{\min}(V) > 0$, we have

$$(\theta_1 - \theta_2)'(G(\theta_1) - G(\theta_2)) \geq (\theta_1 - \theta_2)'(\kappa V)(\theta_1 - \theta_2) > 0$$

for any $\theta_1 \neq \theta_2$. Hence, $G(\theta)$ is an injection from $\mathbb{R}^d$ to $\mathbb{R}^d$, and so G^{-1} is a well-defined function. Consequently, (15) has a unique solution $\hat{\theta} = G^{-1}(Z)$.

Let us consider an η-neighborhood of θ^*, $\mathbb{B}_\eta := \{\theta \;:\; \|\theta - \theta^*\| \leq \eta\}$, where $\eta > 0$ is a constant that will be specified later. Note that $\mathcal{B}_\eta$ is a convex set, thus $\bar{\theta} \in \mathcal{B}_\eta$ as long as $\theta_1, \theta_2 \in \mathbb{B}_\eta$. Define $\kappa_\eta := \inf_{\theta \in \mathbb{B}_\eta} \dot{\mu}(x'\theta) > 0$. From (17), for any $\theta \in \mathcal{B}_\eta$,

$$\begin{aligned} \|G(\theta)\|_{V^{-1}}^2 &= \|G(\theta) - G(\theta^*)\|_{V^{-1}}^2 \\ &= (\theta - \theta^*)' F(\bar{\theta}) V^{-1} F(\bar{\theta})(\theta - \theta^*) \\ &\geq \kappa_\eta^2 \lambda_{\min}(V) \left\|\theta - \theta^*\right\|^2, \end{aligned}$$

where the last inequality is due to the fact that $F(\bar{\theta}) \succeq \kappa_\eta V$.

On the other hand, Lemma A of Chen et al. (1999) implies that

$$\left\{\theta \;:\; \|G(\theta)\|_{V^{-1}} \leq \kappa_\eta \eta \sqrt{\lambda_{\min}(V)}\right\} \subset \mathcal{B}_\eta\,.$$

Now it remains to upper bound $\|Z\|_{V^{-1}} = \left\|G(\hat{\theta})\right\|_{V^{-1}}$ to ensure $\hat{\theta} \in \mathbb{B}_\eta$. To do so, we need the following technical lemma, whose proof is deferred to Section C.

Lemma 7. *Recall σ which is the constant in (2). For any $\delta > 0$, define the following event:*

$$\mathcal{E}_G := \left\{ \|Z\|_{V^{-1}} \leq 4\sigma\sqrt{d + \log(1/\delta)} \right\} .$$

Then, $\mathcal{E}_G$ holds with probability at least $1 - \delta$.

Suppose $\mathcal{E}_G$ holds for the rest of the proof. Then, $\eta \geq \frac{4\sigma}{\kappa_\eta}\sqrt{\frac{d+\log(1/\delta)}{\lambda_{\min}(V)}}$ implies $\left\|\hat{\theta}_t - \theta\right\| \leq \eta$. Since $\kappa = \kappa_1$, we have $\kappa_\eta \geq \kappa$ as long as $\eta \leq 1$. Thus, we have

$$\left\|\hat{\theta} - \theta\right\| \leq \frac{4\sigma}{\kappa}\sqrt{\frac{d + \log(1/\delta)}{\lambda_{\min}(V)}} \leq 1\,, \tag{18}$$

when $\lambda_{\min}(V) \geq 16\sigma^2\left[d + \log(1/\delta)\right]/\kappa^2$.

Step 2: Normality of $\hat{\theta}$. Now, we are ready to precede to prove the normality result. The following assumes $\mathcal{E}_G$ holds (which is high-probability event, according to Lemma 7).

Define $\Delta := \hat{\theta} - \theta^*$. It follows from (17) that there exists a $v \in [0, 1]$ such that

$$Z = G(\hat{\theta}) - G(\theta^*) = (H + E)\Delta\,,$$

where $\tilde{\theta} := v\theta^* + (1 - v)\hat{\theta}$, $H := F(\theta^*) = \sum_{i=1}^n \dot{\mu}(X_i'\theta^*)X_iX_i'$ and $E := F(\tilde{\theta}) - F(\theta^*)$. Intuitively, when $\hat{\theta}$ and θ^* are close, elements in E are small. By the mean value theorem,

$$E = \sum_{i=1}^n \left(\dot{\mu}(X_i'\tilde{\theta}) - \dot{\mu}(X_i'\theta^*)\right) X_iX_i' = \sum_{i=1}^n \ddot{\mu}(r_i)X_i'\Delta X_iX_i'$$

for some $r_i \in \mathbb{R}$. Since $\ddot{\mu} \leq M_\mu$ and $v \in [0, 1]$, for any $x \in \mathbb{R}^d \setminus \{\mathbf{0}\}$, we have

$$\begin{aligned} x'H^{-1/2}EH^{-1/2}x &= (1 - v)\sum_{i=1}^n \ddot{\mu}(r_i)X_i'\Delta\left\|x'H^{-1/2}X_i\right\|^2 \\ &\leq \sum_{i=1}^n M_\mu \|X_i\|\,\|\Delta\|\left\|x'H^{-1/2}X_i\right\|^2 \\ &\leq M_\mu \|\Delta\|\left(x'H^{-1/2}\left(\sum_{i=1}^n X_iX_i'\right)H^{-1/2}x\right) \\ &\leq \frac{M_\mu}{\kappa}\|\Delta\|\,\|x\|^2\,, \end{aligned}$$

where we have used the assumption that $\|X_i\| \leq 1$ for the second inequality. Therefore,

$$\left\|H^{-1/2}EH^{-1/2}\right\| \leq \frac{M_\mu}{\kappa}\|\Delta\| \leq \frac{4M_\mu\sigma}{\kappa^2}\sqrt{\frac{d + \log(1/\delta)}{\lambda_{\min}(V)}}\,. \tag{19}$$

When $\lambda_{\min}(V) \geq 64M_\mu^2\sigma^2(d + \log(1/\delta))/\kappa^4$, we have

$$\left\|H^{-1/2}EH^{-1/2}\right\| \leq 1/2\,. \tag{20}$$

Now we are ready to prove the theorem. For any $x \in \mathbb{R}^d$,

$$x'(\hat{\theta} - \theta^*) \;=\; x'(H + E)^{-1}Z \;=\; x'H^{-1}Z - x'H^{-1}E(H + E)^{-1}Z\,. \tag{21}$$

Note that the matrix $(H + E)$ is nonsingular, so its inversion exists.

For the first term, $\{\epsilon_i\}$ are sub-Gaussian random variables with sub-Gaussian parameter σ. Define

$$D := [X_1, X_2, \ldots, X_n]' \in \mathbb{R}^{n\times d}$$

to be the design matrix. Hoeffding inequality gives

$$\mathbb{P}\{|x'H^{-1}Z| \geq t\} \leq 2\exp\left\{-\frac{t^2}{2\sigma^2 \left\|x'H^{-1}D'\right\|^2}\right\}. \tag{22}$$

Since $H \succeq \kappa V = \kappa D'D$, we have

$$\left\|x'H^{-1}D'\right\|^2 = x'H^{-1}D'DH^{-1}x \leq \frac{1}{\kappa^2}x'V^{-1}x = \frac{1}{\kappa^2}\left\|x\right\|_{V^{-1}}^2 \,,$$

so (22) implies

$$\mathbb{P}\{|x'H^{-1}Z| \geq t\} \leq 2\exp\left\{-\frac{t^2\kappa^2}{2\sigma^2 \left\|x\right\|_{V^{-1}}^2}\right\}.$$

Let the right-hand side be 2δ and solve for t, we obtain that with probability at least $1-2\delta$,

$$|x'H^{-1}Z| \leq \frac{\sqrt{2}\sigma}{\kappa}\sqrt{\log(1/\delta)}\left\|x\right\|_{V^{-1}}\,. \tag{23}$$

For the second term,

$$\begin{aligned}
|x'H^{-1}E(H+E)^{-1}Z| &\leq \left\|x\right\|_{H^{-1}}\left\|H^{-1/2}E(H+E)^{-1}Z\right\| \\
&\leq \left\|x\right\|_{H^{-1}}\left\|H^{-1/2}E(H+E)^{-1}H^{1/2}\right\|\left\|Z\right\|_{H^{-1}} \\
&\leq \frac{1}{\kappa}\left\|x\right\|_{V^{-1}}\left\|H^{-1/2}E(H+E)^{-1}H^{1/2}\right\|\left\|Z\right\|_{V^{-1}}\,,
\end{aligned} \tag{24}$$

where the last inequality is due to the fact that $H \succeq \kappa V$. Since $(H+E)^{-1} = H^{-1} - H^{-1}E(H+E)^{-1}$, we have

$$\begin{aligned}
\left\|H^{-1/2}E(H+E)^{-1}H^{1/2}\right\| &= \left\|H^{-1/2}E\left(H^{-1} - H^{-1}E(H+E)^{-1}\right)H^{1/2}\right\| \\
&= \left\|H^{-1/2}EH^{-1/2} + H^{-1/2}EH^{-1}E(H+E)^{-1}H^{1/2}\right\| \\
&\leq \left\|H^{-1/2}EH^{-1/2}\right\| + \left\|H^{-1/2}EH^{-1/2}\right\|\left\|H^{-1/2}E(H+E)^{-1}H^{1/2}\right\|.
\end{aligned}$$

By solving this inequality, we get

$$\left\|H^{-1/2}E(H+E)^{-1}H^{1/2}\right\| \leq \frac{\left\|H^{-1/2}EH^{-1/2}\right\|}{1-\left\|H^{-1/2}EH^{-1/2}\right\|} \leq 2\left\|H^{-1/2}EH^{-1/2}\right\| \leq \frac{8M_\mu\sigma}{\kappa^2}\sqrt{\frac{d+\log(1/\delta)}{\lambda_{\min}(V)}}\,,$$

where we have used (20) and (19) in the second and third inequalities, respectively. Combining it with (24) and the bound in $\mathcal{E}_G$, we have

$$|x'H^{-1}E(H+E)^{-1}Z| \leq \frac{32M_\mu\sigma^2}{\kappa^3}\frac{d+\log(1/\delta)}{\sqrt{\lambda_{\min}(V)}}\left\|x\right\|_{V^{-1}}. \tag{25}$$

From (21), (23) and (25), one can see that (5) holds as long as the lower bound (4) for $\lambda_{\min}(V)$ holds. Finally, an application of a union bound on two small-probability events (given in Lemma 7 and (23), respectively) asserts that (5) holds with probability at least $1-3\delta$.

B. Proof of Proposition 1

In the following, for simplicity, we will drop the subscript n when there is no ambiguity. Therefore, V_n is denoted V and so on.

Let X be a random vector drawn from the distribution ν. Define $Z := \Sigma^{-1/2}X$. Then Z is isotropic, namely, $\mathbb{E}[ZZ'] = \mathbf{I}_d$. Define $U = \sum_{t=1}^n Z_t Z_t' = \Sigma^{-1/2}V\Sigma^{-1/2}$. From Lemma 1, we have that, for any t, with probability at least $1-2\exp(-C_2t^2)$,

$$\lambda_{\min}(U) \geq n - C_1\sigma^2\sqrt{nd} - \sigma^2 t\sqrt{n}\,.$$

where σ is the sub-Gaussian parameter of Z, which is upper-bounded by $\left\|\Sigma^{-1/2}\right\| = \lambda_{\min}^{-1/2}(\Sigma)$ (see, e.g., Vershynin (2012)). We thus can rewrite the above inequality (which holds with probability $1-\delta$ as

$$\lambda_{\min}(U) \geq n - \lambda_{\min}^{-1}(\Sigma)\left(C_1\sigma^2\sqrt{nd} + t\sqrt{n}\right)\,.$$

We now bound the minimum eigenvalue of V, as follows:

$$\begin{aligned}
\lambda_{\min}(V) &= \min_{x\in\mathbb{B}^d} x'Vx \\
&= \min_{x\in\mathbb{B}^d} x'\Sigma^{1/2}U\Sigma^{1/2}x \\
&\geq \lambda_{\min}(U)\min_{x\in\mathbb{B}^d} x'\Sigma x \\
&= \lambda_{\min}(U)\lambda_{\min}(\Sigma) \\
&\geq \lambda_{\min}(\Sigma)\left(n - \lambda_{\min}^{-1}(\Sigma)(C_1\sigma^2\sqrt{nd} + t\sqrt{n})\right) \\
&= \lambda_{\min}(\Sigma)n - C_1\sqrt{nd} - C_2\sqrt{n\log(1/\delta)}\,.
\end{aligned}$$

Finally, it can be verified (Lemma 9) that the last expression above is no less than B as long as

$$n \geq \left(\frac{C_1\sqrt{d} + C_2\sqrt{\log(1/\delta)}}{\lambda_{\min}(\Sigma)}\right)^2 + \frac{2B}{\lambda_{\min}(\Sigma)}\,,$$

finishing the proof.

C. Technical Lemmas and Proofs

C.1. Proof of Lemma 7

Noting that

$$\|Z\|_{V^{-1}} = \|V^{-1/2}Z\|_2 = \sup_{\|a\|_2\leq 1}\langle a, V^{-1/2}Z\rangle,$$

let $\hat{\mathbb{B}}$ be a $1/2$-net of the unit ball $\mathbb{B}^d$. Then $|\hat{\mathbb{B}}| \leq 6^d$ (Pollard, 1990, Lemma 4.1), and for any $x \in \mathbb{B}^d$, there is a $\hat{x} \in \hat{\mathbb{B}}$ such that $\|x-\hat{x}\| \leq 1/2$. Consequently,

$$\begin{aligned}
\langle x, V^{-1/2}Z\rangle &= \langle \hat{x}, V^{-1/2}Z\rangle + \langle x-\hat{x}, V^{-1/2}Z\rangle \\
&= \langle \hat{x}, V^{-1/2}Z\rangle + \|x-\hat{x}\|\,\langle\frac{x-\hat{x}}{\|x-\hat{x}\|}, V^{-1/2}Z\rangle \\
&\leq \langle \hat{x}, V^{-1/2}Z\rangle + \frac{1}{2}\sup_{z\in\mathbb{B}^d}\langle z, V^{-1/2}Z\rangle.
\end{aligned}$$

Taking supremum on both sides, we get

$$\sup_{x\in\mathbb{B}^d}\langle x, V^{-1/2}Z\rangle \leq 2\max_{\hat{x}\in\hat{\mathbb{B}}}\langle\hat{x}, V^{-1/2}Z\rangle\,.$$

Then a union bound argument implies

$$\begin{aligned}
\mathbb{P}\left\{\|Z\|_{V^{-1}} > t\right\} &\leq \mathbb{P}\left\{\max_{\hat{x}\in\hat{\mathbb{B}}}\langle \hat{x}, V^{-1/2}Z\rangle > t/2\right\} \\
&\leq \sum_{\hat{x}\in\hat{\mathbb{B}}} \mathbb{P}\left\{\langle \hat{x}, V^{-1/2}Z\rangle > t/2\right\} \\
&\leq \sum_{\hat{x}\in\hat{\mathbb{B}}} \exp\left\{-\frac{t^2}{8\sigma^2\left\|\hat{x}'V^{-1/2}X'\right\|^2}\right\} \\
&\leq \exp\left\{-t^2/(8\sigma^2) + d\log 6\right\},
\end{aligned}$$

where we have used Hoeffding's inequality for the third inequality and $|\hat{\mathbb{B}}| \leq 6^d$ for the last inequality. A choice of $t = 4\sigma\sqrt{d + \log(1/\delta)}$ completes the proof.

C.2. Proof of Lemma 2

By Abbasi-Yadkori et al. (2011, Lemma 11), we have

$$\sum_{t=m+1}^{m+n} \|X_t\|^2_{V_t^{-1}} \leq 2\log\frac{\det V_{m+n+1}}{\det V_{m+1}} \leq 2d\log\left(\frac{\operatorname{tr}(V_{m+1}) + n}{d}\right) - 2\log\det V_{m+1}\,.$$

Note that $\operatorname{tr}(V_{m+1}) = \sum_{t=1}^m \operatorname{tr}(X_tX_t') = \sum_{t=1}^m \|X_t\|^2 \leq m$ and that $\det V_{m+1} = \prod_{i=1}^d \lambda_i \geq \lambda_{\min}^d(V_{m+1}) \geq 1$, where $\{\lambda_i\}$ are the eigenvalues of V_{m+1}. Applying Cauchy-Schwartz inequality yields

$$\sum_{t=m+1}^{m+n} \|X_t\|_{V_t^{-1}} \leq \sqrt{n\sum_{t=m+1}^{m+n}\|X_t\|^2_{V_t^{-1}}} \leq \sqrt{2nd\log\left(\frac{n+m}{d}\right)}\,.$$

C.3. Proof of Lemma 3

Define $G_t(\theta) = \sum_{i=1}^{t-1}(\mu(X_i'\theta) - \mu(X_i'\theta^*))X_i$ and $Z_t = \sum_{i=1}^{t-1}\epsilon_i X_i$. Following the same argument as in the proof of Theorem 1, we have $G_t(\hat{\theta}_t) = Z_t$ and

$$\|G_t(\theta)\|^2_{V_t^{-1}} \geq \kappa^2\|\theta - \theta^*\|^2_{V_t} \tag{26}$$

for any $\theta \in \{\theta \ : \ \|\theta - \theta^*\| \leq 1\}$. Combining (26) with the following lemma and the equality $Z_t = G_t(\hat{\theta}_t)$ completes the proof.

Lemma 8. *Suppose there is an integer m such that $\lambda_{\min}(V_m) \geq 1$, then for any $\delta \in (0,1)$, with probability at least $1-\delta$, for all $t > m$,*

$$\|Z_t\|^2_{V_t^{-1}} \leq 4\sigma^2\left(\frac{d}{2}\log(1 + 2t/d) + \log(1/\delta)\right)\,.$$

Proof. For convenience, fix t such that $t > m$, and denote V_t and Z_t by V and Z, respectively. Furthermore, define $\bar{V} := V + \lambda I$ and let $\mathbf{1}$ be the vector of all 1s. It is easy to observe that

$$\|Z\|^2_{V^{-1}} = \|Z\|^2_{\bar{V}^{-1}} + Z'(V^{-1} - \bar{V}^{-1})Z\,. \tag{27}$$

We start with bounding the second term. The ShermanMorrison formula gives

$$\bar{V}^{-1} = V^{-1} - \frac{\lambda V^{-2}}{1 + \lambda\mathbf{1}'V^{-1}\mathbf{1}}\,.$$

Since $\mathbf{1}'V^{-1}\mathbf{1} \geq 0$, the above implies that

$$\begin{aligned}
0 &\leq Z'(V^{-1} - \bar{V}^{-1})Z \\
&\leq \lambda Z'V^{-2}Z \\
&\leq \lambda\left\|V^{-1}\right\|\|Z\|^2_{V^{-1}} \\
&= \frac{\lambda}{\lambda_{\min}(V)}\|Z\|^2_{V^{-1}}\,.
\end{aligned}$$

Since $\lambda_{\min}(V) \geq \lambda_{\min}(V_m) \geq 1$, we now have

$$0 \leq Z'(V^{-1} - \bar{V}^{-1})Z \leq \lambda \|Z\|^2_{V^{-1}} .$$

The above inequality together with (27) implies that

$$\|Z\|^2_{V^{-1}} \leq (1-\lambda)^{-1} \|Z\|^2_{\bar{V}^{-1}} .$$

The proof can be finished by applying Theorem 1 and Lemma 10 from Abbasi-Yadkori et al. (2011) to bound $\|Z\|^2_{\bar{V}^{-1}}$, using $\lambda = 1/2$. □

C.4. Proof of Lemma 6

We will prove the first part of the lemma by induction. It is easy to check the lemma holds for $s = 1$. Suppose we have $a_t^* \in A_s$ and we want to prove $a_t^* \in A_{s+1}$. Since the algorithm proceeds to stage $s+1$, we know from step 2b that

$$|m_{t,a}^{(s)} - x_{t,a}'\theta^*| \leq w_{t,a}^{(s)} \leq 2^{-s}$$

for all $a \in A_s$. Specially, it holds for $a = a_t^*$ because $a_t^* \in A_s$ by our induction step. Then the optimality of a_t^* implies

$$m_{t,a_t^*}^{(s)} \geq x_{t,a_t^*}'\theta^* - 2^{-s} \geq x_{t,a}'\theta^* - 2^{-s} \geq m_{t,a}^{(s)} - 2 \cdot 2^{-s}$$

for all $a \in A_s$. Thus we have $a_t^* \in A_{s+1}$ according to step 2d.

Suppose a_t is selected at stage s_t in step 2b. If $s_t = 1$, obviously the lemma holds because $0 \leq \mu(x) \leq 1$ for all x. If $s_t > 1$, since we have proved $a_t^* \in A_{s_t}$, again step 2b at stage $s_t - 1$ implies

$$|m_{t,a}^{(s_t-1)} - x_{t,a}'\theta^*| \leq 2^{-s_t+1}$$

for $a = a_t$ and $a = a_t^*$. Step 2d at stage $s_t - 1$ implies

$$m_{t,a_t^*}^{(s_t-1)} - m_{t,a_t}^{(s_t-1)} \leq 2 \cdot 2^{-s_t+1} .$$

Combining above two inequalities, we get

$$x_{t,a_t}'\theta^* \geq m_{t,a_t}^{(s_t-1)} - 2^{-s_t+1} \geq m_{t,a_t^*}^{(s_t-1)} - 3 \cdot 2^{-s_t+1} \geq x_{t,a_t^*}'\theta^* - 4 \cdot 2^{-s_t+1} .$$

When a_t is selected in step 2c, since $m_{t,a_t}^{(s_t)} \geq m_{t,a_t^*}^{(s_t)}$, we have

$$x_{t,a_t}'\theta^* \geq m_{t,a_t}^{(s_t)} - 1/\sqrt{T} \geq m_{t,a_t^*}^{(s_t)} - 1/\sqrt{T} \geq x_{t,a_t^*}'\theta^* - 2/\sqrt{T} .$$

Using the fact that $\mu(x_1) - \mu(x_2) \leq L_\mu(x_1 - x_2)$ for $x_1 \geq x_2$, we will get the desired result.

C.5. Proof of Lemma 9

Lemma 9. *Let a and b be two positive constants. If $m \geq a^2 + 2b$, then $m - a\sqrt{m} - b \geq 0$.*

Proof. The function $t \mapsto t^2 - at - b$ is monotonically increasing for $t \geq a/2$. Since $m \geq a^2 + 2b$, we have $\sqrt{m} \geq a/2$, so

$$\begin{aligned} m - a\sqrt{m} - b &\geq a^2 + 2b - a\sqrt{a^2 + 2b} - b \\ &\geq a^2 + b - a\sqrt{a^2 + 2b + b^2/a^2} \\ &= a^2 + b - a\sqrt{(a + b/a)^2} \\ &= a^2 + b - a(a + b/a) \\ &= 0 . \end{aligned}$$

□