APPENDIX: On Blackbox Backpropagation and Jacobian Sensing

In the Appendix we prove theoretical results from the main body of the paper and add technical discussion regarding presented coloring algorithm. We denote by $\|\cdot\|$ the L_2-norm of the vector.

We will show a strengthened version of Theorem 3.3 proving that there are many more classes of distributions than: Gaussian, Poissonian and bounded with positive variance for the choice of the measurement directions. We need to introduce the following definition.

Definition 4.1 (regular distributions)**.** *We say that a random variable ϕ is $A(n)$-regular for some function A if there exists a constant $\tau > 0$ such that for any $x \in \mathbb{R}_+$ and n large enough the following holds:*

$$\min_{s \in \mathbb{R}_+} f_{n,\phi}(s) < sx - (2+\tau)\log(x), \tag{4}$$

where $f_{n,\phi}(s) = n \max_{\xi:|\xi| \leq 2sA(n)} \log \mathbb{E}[e^{\xi(\phi - \mathbb{E}[\phi])}]$.

4.1 Proof of Theorem 3.3

Before we give the proof of the result, we prove the following lemma that turns out to play crucial role in the entire analysis.

Lemma 4.2 (random measurement directions lemma)**.** *Assume that each $d_j^i \sim \phi$ is chosen independently, where*

- *ϕ is Gaussian, Poissonian or bounded with finite positive variance or*
- *ϕ is $\frac{M(n)}{D(n)}$-regular for some functions $M(n), D(n) > 0$.*

Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$ and assume that $\max_{i,j} |A_{i,j}|, \max_{i,j} |B_{i,j}| \leq M(n)$ (this condition is required only if we consider regular distributions that are not: Gaussian, Poissonian or bounded with positive variance). If furthermore $\|\mathbf{A} - \mathbf{B}\|_F > D(n)$, where $\|\cdot\|_F$ stands for the Frobenius norm and we take $\eta(n) = o(D(n)\sqrt{Var(\phi)})$, then there exists $\delta > 0$ such that for n large enough:

$$\mathbb{P}[\|(\mathbf{A} - \mathbf{B})\mathbf{d}\| > 2\eta(n) + 1] > \delta,$$

where the probability is chosen with respect to the random choices of the random direction $\mathbf{d} = (d_1, ..., d_n)^\top$.

Proof. Denote $\mathbf{V}_i = (\mathbf{A}^i - \mathbf{B}^i)d_i$ for $i = 1, ..., n$, where $\mathbf{X}^i$ for $\mathbf{X} \in \mathbb{R}^{m \times n}$ stands for the ith column of $\mathbf{X}$. Denote $\mathbf{V} = \sum_{i=1}^n \mathbf{V}_i$. Note that $\mathbf{V} = (\mathbf{A} - \mathbf{B})\mathbf{d}$ and $\mathbf{V} \in \mathbb{R}^m$. Denote $P = \|\mathbf{V} - \mathbb{E}[\mathbf{V}]\|_2^2$, where $\mathbb{E}[\mathbf{V}] \in \mathbb{R}^m$ stands for the mean vector of the vector random variable $\mathbf{V}$. We need to prove that

$$\mathbb{P}[\|\mathbf{V}\| > 2\eta(n) + 1] > \delta. \tag{5}$$

Take a constant $T > 0$ (its exact value will be determined later).

We will consider the following cases.

Case 1: $\|\mathbb{E}[\mathbf{V}]\| \leq T\sqrt{\mathbb{E}[P]}$

In that setting it suffices to show that $\mathbb{P}[\|\mathbf{V} - \mathbb{E}[\mathbf{V}]\| > 2T\sqrt{\mathbb{E}[P]}] > \delta$ for δ from the statement of the lemma. Indeed, from the latter inequality and the triangle inequality, we get:

$$\mathbb{P}[\|\mathbf{V}\| > T\sqrt{\mathbb{E}[P]}] > \delta. \tag{6}$$

Thus our claims is correct if $T\sqrt{\mathbb{E}[P]} \geq 2\eta(n) + 1$. Denote $\mathbf{C} = \mathbf{A} - \mathbf{B}$. We have $\mathbf{V} - \mathbb{E}[\mathbf{V}] = \sum_{i=1}^n \mathbf{C}^i(d_i - \mathbb{E}[d_i])$. Thus, from the independence of $d_i s$ we get: $\mathbb{E}[\|\sum_{i=1}^n \mathbf{C}^i(d_i - \mathbb{E}[d_i])\|^2] = \|\mathbf{C}\|_F^2 Var(\phi)$, thus $\sqrt{\mathbb{E}[P]} = \|\mathbf{C}\|_F\sqrt{Var(\phi)}$. Therefore, since $\eta(n) = o(D(n)\sqrt{Var(\phi)})$, the claim is correct.

Now we partition the probabilistic space into three regions:

- $\mathcal{R}_1 = \{P < \alpha\mathbb{E}[P]\}$
- $\mathcal{R}_2 = \{\alpha\mathbb{E}[P] \leq P \leq \beta\mathbb{E}[P]\}$
- $\mathcal{R}_3 = \{P > \beta\mathbb{E}[P]\}$,

where again as in the case of T, constants α and β will be given explicitly later.

The following is trivially true:

$$\mathbb{E}[P] = \mathbb{P}[\mathcal{R}_1]\mathbb{E}(P|\mathcal{R}_1) + \mathbb{P}[\mathcal{R}_2]\mathbb{E}(P|\mathcal{R}_2) + \mathbb{P}[\mathcal{R}_3]\mathbb{E}(P|\mathcal{R}_3). \tag{7}$$

We obtain:

$$\mathbb{E}[P] \leq \alpha \mathbb{E}[P] + \mathbb{P}[\mathcal{R}_2] \cdot \beta \mathbb{E}[P] + \mathbb{P}[\mathcal{R}_3]\mathbb{E}(P|\mathcal{R}_3). \tag{8}$$

Our next goal is to show that for a sufficiently large constant β, we obtain:

$$\mathbb{P}[\mathcal{R}_3]\mathbb{E}(P|\mathcal{R}_3) \leq \alpha \mathbb{E}[P]. \tag{9}$$

To see, why it completes our analysis in Case 1, notice that from Equation 8 we then know that:

$$\mathbb{P}[\mathcal{R}_2] \geq \frac{1-2\alpha}{\beta} > 0 \tag{10}$$

for $\alpha < \frac{1}{2}$. Take $T = \sqrt{\frac{\alpha}{4}}$. We obtain:

$$\mathbb{P}[\|\mathbf{V} - \mathbb{E}[\mathbf{V}]\| > 2T\sqrt{\mathbb{E}[P]}] = \mathbb{P}[P > \alpha \mathbb{E}[P]] \geq \mathbb{P}[\mathcal{R}_2] \geq \frac{1-2\alpha}{\beta} \tag{11}$$

and thus for $\delta = \frac{1-2\alpha}{\beta}$ the proof is completed.

Thus it remains to prove that $\mathbb{P}[\mathcal{R}_3]\mathbb{E}(P|\mathcal{R}_3) \leq \alpha \mathbb{E}[P]$ for constant β large enough. Therefore it suffices to prove that

$$\mathbb{E}[\frac{P}{\mathbb{E}[P]} I\{\frac{P}{\mathbb{E}[P]} > \beta\}] < \alpha, \tag{12}$$

where $I(\cdot)$ stands for the indicator random variable. Note that we have:

$$\mathbb{E}[\frac{P}{\mathbb{E}[P]} I\{\frac{P}{\mathbb{E}[P]} > \beta\}] = \beta \mathbb{P}[\frac{P}{\mathbb{E}[P]} > \beta] + \int_{\beta}^{\infty} \mathbb{P}[\frac{P}{\mathbb{E}[P]} > y] dy. \tag{13}$$

Note that for any $c > 0$

$$\mathbb{P}[\frac{P}{\mathbb{E}[P]} > c] = \mathbb{P}[P > c\mathbb{E}[P]] = \mathbb{P}[\sum_{j=1}^{m} P_j > c\mathbb{E}[P]], \tag{14}$$

where $P_j = (\sum_{i=1}^{n} C_j^i (d_i - \mathbb{E}[d_i]))^2$ and C_j^i stands for the jth element of $\mathbf{C}^i \in \mathbb{R}^m$.

Therefore, we obtain:

$$\mathbb{P}[\frac{P}{\mathbb{E}[P]} > c] \leq \mathbb{P}[\exists_{j \in \{1,...,m\}} P_j \geq \frac{c\mathbb{E}[P]}{m}] \leq m \cdot \sup_{j \in \{1,...,m\}} \mathbb{P}[P_j \geq \frac{c\mathbb{E}[P]}{m}], \tag{15}$$

where the last inequality comes from the union bound. Now fix some $j \in \{1, ..., m\}$. Our goal will be to find an upper bound on $\mathbb{P}[P_j \geq \frac{c\mathbb{E}[P]}{m}]$.

Note that

$$\mathbb{P}[P_j > \frac{c\mathbb{E}[P]}{m}] = \mathbb{P}[|\sum_{i=1}^{n} C_j^i (d_i - \mathbb{E}[d_i])| > \sqrt{\frac{c}{m}}\sqrt{\mathbb{E}[P]}] \leq \mathbb{P}[|\sum_{i=1}^{n} C_j^i (d_i - \mathbb{E}[d_i])| > D(n)\sqrt{\frac{cVar(\phi)}{m}}], \tag{16}$$

where the last inequality comes from the derived earlier formula for $\mathbb{E}[P]$ and the definition of function $D(n)$.

We conclude that

$$\mathbb{P}[P_j > \frac{c\mathbb{E}[P]}{m}] \leq \mathbb{P}[\sum_{i=1}^{n} X_i > \sqrt{\frac{cVar(\phi)}{m}}] + \mathbb{P}[\sum_{i=1}^{n} (-X_i) > \sqrt{\frac{cVar(\phi)}{m}}], \tag{17}$$

where $X_i = \frac{C_j^i}{D(n)}(d_i - \mathbb{E}[d_i])$. Notice that random variables X_i are independent since d_is are independent, and furthermore, $\mathbb{E}[X_i] = 0$ for $i = 1, ..., n$.

Assume first that each d_i is bounded, i.e. there exists some constant $U > 0$ such that $|d_i| \leq U$ for $i = 1, ..., n$. Then we have: $|X_i| \leq \frac{2U|C_j^i|}{D(n)}$. We will use Azuma's inequality:

Lemma 4.3 (Azuma's Inequality)**.** *If $X = \sum_{i=1}^{n} X_i$, where X_is are independent with zero mean and furthermore $|X_i| \leq c_i$ then the following holds:*

$$\mathbb{P}[|X| > t] \leq 2e^{-\frac{t^2}{\sum_{i=1}^{n} c_i^2}}. \tag{18}$$

Thus we obtain:

$$\mathbb{P}[\sum_{i=1}^{n} X_i > \sqrt{\frac{cVar(\phi)}{m}}] + \mathbb{P}[\sum_{i=1}^{n}(-X_i) > \sqrt{\frac{cVar(\phi)}{m}}] \leq 2e^{-\frac{\frac{cVar(\phi)}{m}}{4U^2 \sum_{i=1}^{n} \frac{|C_j^i|^2}{D^2(n)}}}. \tag{19}$$

Note that $\sum_{i=1}^{n} \frac{|C_j^i|^2}{D^2(n)} \leq 1$, thus we get:

$$\mathbb{P}[\sum_{i=1}^{n} X_i > \sqrt{\frac{cVar(\phi)}{m}}] + \mathbb{P}[\sum_{i=1}^{n}(-X_i) > \sqrt{\frac{cVar(\phi)}{m}}] \leq 2e^{-\frac{cVar(\phi)}{4mU^2}}. \tag{20}$$

Thus we get

$$\mathbb{P}[\frac{P}{\mathbb{E}[P]} > c] \leq 2e^{-\frac{cVar(\phi)}{4mU^2}}. \tag{21}$$

Plugging in this formula for $c = \beta$ and $c = y$ to Equation 13, we conclude that for β large enough the RHS of the equation is arbitrarily close to 0 and that completes Case 1 of the proof for bounded d_is.

Now let us assume that d_is are Gaussian or Poissonian with constant variance. But then each X_i is also Gaussian/Poissonian with constant variance, where the latter comes from the fact that $\sum_{i=1}^{n} \frac{|C_j^i|^2}{D^2(n)} \leq 1$ and the fact that the sum of independent Gaussian/Poissonian random variables is a Gaussian/Poissonian random variable with variance which is equal to the sum of variances of the random variables in the sum. Therefore, we can repeat the second part of the analysis of the bounded case (using standard upper bounds on tails of the Gaussian and Poissonian distributions) and complete the proof for Case 1.

To obtain strong bounds for a wide variety of other random variables d_i for probabilities from Inequality 17, we will apply Chernoff-Cramer method.

The following is true

Theorem 4.4 (Chernoff-Cramer method)**.** *Let X be a centered random variable such that $M_X(s) < +\infty$ on $s \in (-s_0, s_0)$ for some $s_0 > 0$, where $M_X(s) = \mathbb{E}[e^{sX}]$. Then for any $\beta > 0$ the following holds:*

$$\mathbb{P}[X \geq \beta] \leq e^{-\Psi_X^*(\beta)}, \tag{22}$$

where

$$\Psi_X^*(\beta) = \sup_{s \in \mathbb{R}_+} (s\beta - \log(M_X(s))). \tag{23}$$

Applying the above inequality to our random variables X_i and taking $d = \frac{cVar(\phi)}{m}$, we obtain:

$$\mathbb{P}[\sum_{i=1}^{n} X_i > \sqrt{d}] \leq e^{-\gamma_\phi^1(\sqrt{d})}, \tag{24}$$

and

$$\mathbb{P}[\sum_{i=1}^{n}(-X_i) > \sqrt{d}] \leq e^{-\gamma_\phi^2(\sqrt{d})}, \tag{25}$$

where

$$\gamma_\phi^1(x) = \sup_{s \in \mathbb{R}_+} (sx - \sum_{i=1}^{n} \log(\mathbb{E}[e^{s\frac{C_j^i}{D(n)}(d_i - \mathbb{E}[d_i])}])) \tag{26}$$

and

$$\gamma_\phi^2(x) = \sup_{s \in \mathbb{R}_+} (sx - \sum_{i=1}^{n} \log(\mathbb{E}[e^{-s\frac{C_j^i}{D(n)}(d_i - \mathbb{E}[d_i])}])), \tag{27}$$

where we use the fact that d_ia are independent random variables. Now note that if d_i is taken from the class of $\frac{M(n)}{D(n)}$-regular distributions then $\gamma_\phi^1, \gamma_\phi^2 \geq (2+\chi)\log(x)$ for n, x large enough and some constant $\chi > 0$. Therefore we obtain:

$$\mathbb{P}[\sum_{i=1}^{n} X_i > \sqrt{d}] \leq \frac{1}{d^{1+\frac{\chi}{2}}} \tag{28}$$

and

$$\mathbb{P}[\sum_{i=1}^{n}(-X_i) > \sqrt{d}] \leq \frac{1}{d^{1+\frac{\chi}{2}}}. \tag{29}$$

Then we continue as before and from Equation 13 we obtain:

$$\mathbb{E}[\frac{P}{\mathbb{E}[P]} I\{\frac{P}{\mathbb{E}[P]} > \beta\}] \leq 2\beta \frac{1}{(\frac{\beta Var(\phi)}{m})^{1+\frac{\chi}{2}}} + 2\int_{\beta}^{\infty} \frac{1}{(\frac{yVar(\phi)}{m})^{1+\frac{\chi}{2}}} dy. \tag{30}$$

Both expressions on the RHS of the inequality above are arbitrarily close to 0 for β large enough since the integral $\int_1^{\infty} \frac{dy}{y^{1+\frac{\chi}{2}}}$ is finite. Thus Case 1 of the proof is completed, provided that the explicit value of α is given (notice that α determines T).

Now let us consider the remaining case.

Case 2: $\|\mathbb{E}[\mathbf{V}]\| > T\sqrt{\mathbb{E}[P]}$

Notice that if $\|\mathbb{E}[\mathbf{V}]\| > T\sqrt{\mathbb{E}[P]}$ then there exists $j \in \{1, ..., m\}$ such that $|(\mathbb{E}[\mathbf{V}])_j| \geq \frac{T\sqrt{\mathbb{E}[P]}}{\sqrt{m}}$.

Without loss of generality, we will assume that $|(\mathbb{E}[\mathbf{V}])_1| \geq \frac{T\sqrt{\mathbb{E}[P]}}{\sqrt{m}}$. Analysis of the other cases is exactly the same. Without loss of generality, we can assume further that $(\mathbb{E}[\mathbf{V}])_1 \geq \frac{T\sqrt{\mathbb{E}[P]}}{\sqrt{m}}$.

As before, we observe that it suffices to show that

$$\mathbb{P}[\|\mathbf{V}\| > \tau\sqrt{\mathbb{E}[P]}] > \delta \tag{31}$$

for some constants $\delta, \tau > 0$.

Note that

$$\mathbb{P}[\|\mathbf{V}\| \leq \tau\sqrt{\mathbb{E}[P]}] \leq \mathbb{P}[|V_1| \leq \tau\sqrt{\mathbb{E}[P]}] \leq \mathbb{P}[V_1 - \mathbb{E}[V_1] \leq -(\frac{T}{\sqrt{m}} - \tau)\sqrt{\mathbb{E}[P]}]. \tag{32}$$

Thus we conclude that

$$\mathbb{P}[\|\mathbf{V}\| > \tau\sqrt{\mathbb{E}[P]}] \geq 1 - \mathbb{P}[V_1 - \mathbb{E}[V_1] \leq -(\frac{T}{\sqrt{m}} - \tau)\sqrt{\mathbb{E}[P]}]. \tag{33}$$

Note that

$$\mathbb{P}[V_1 - \mathbb{E}[V_1] \leq -(\frac{T}{\sqrt{m}} - \tau)\sqrt{\mathbb{E}[P]}] = \mathbb{P}[\sum_{i=1}^{n} C_1^i(d^i - \mathbb{E}[d^i]) \leq -(\frac{T}{\sqrt{m}} - \tau)\sqrt{\mathbb{E}[P]}] \tag{34}$$

Take $\alpha = \frac{1}{3}$ and $T = \sqrt{\frac{\alpha}{4}} = \sqrt{\frac{1}{12}}$. We will take $\lambda = \sqrt{\frac{1}{13}}$ and $\tau = \frac{T}{2\sqrt{m}}$.

Note that if we take: $\sqrt{\frac{c}{m}} = \frac{T}{\sqrt{m}-\tau}$ then we can apply previous upper bounds on the probability from the inequality above. Those for ϕ being Gaussian, Poissonian or bounded are clearly separated from $p = 1$. The one obtained via the Chernoff-Cramer method also is, provided that $Var(\phi) > 48m$, because if this condition is satisfied then $d > 1$, where $d = \frac{cVar(\phi)}{m}$. That completes the proof of the lemma. □

We are ready to prove our main result in this section.

Proof. Define: $A' = d_{\text{int}} \log(\frac{C\sqrt{mn}}{E(n)})$, $B' = m\rho(\mathbf{J}, G_{\text{int}}^{\text{weak}}) \log(\frac{C\sqrt{m\rho(\mathbf{J}, G_{\text{int}}^{\text{weak}}}}{E(m)})$ Consider a solution $\widehat{\mathbf{J}}$ to the proposed LP program. Consider the set of N balls covering the feasibility region of the convex optimization problem that is not defined by the random directions $\mathbf{d}^i$ and such that any element of that region is within a distance $D(n)$ from one of the balls. We call this set a *grids* since its goal is to accurately cover the entire feasibility region.

Assume first that the algorithm explores the sparsity structure of the Jacobian via its weak-intersection graph $G_{\text{int}}^{\text{weak}}$. Notice that the coloring uses $O(\rho(\mathbf{J}, G_{\text{int}}^{\text{weak}}))$ number of colors, thus the compressed problem has this many vector variables (see: proof of Lemma 3.1 for the explanation). Then the intrinsic dimensionality of the space is clearly $m\rho(\mathbf{J}, G_{\text{int}}^{\text{weak}})$ since the number of vector variables is exactly $\rho(\mathbf{J}, G_{\text{int}}^{\text{weak}})$. Thus, from the definition of intrinsic dimensionality, we conclude that one can take $N = B'$. On the other hand, if the algorithm just explores the structure given by $\mathcal{C}$, then one can take $N = A'$ (again, by exploiting intrinsic dimensionality, this time of the set $\mathcal{C}$).

Denote by $\widehat{\mathbf{J}}_{\text{round}}$ the element of the grid that is closest to $\widehat{\mathbf{J}}$ in the Frobenius norm sense. Denote by $\mathbf{J}$ the true Jacobian. Consider some random direction $\mathbf{d} \in \mathbb{R}^n$ and the corresponding measurement $\mathbf{r} \in \mathbb{R}^m$. We have:

$$\|\widehat{\mathbf{J}}_{\text{round}}\mathbf{d} - \mathbf{J}\mathbf{d}\| \leq \|\widehat{\mathbf{J}}_{\text{round}}\mathbf{d} - \widehat{\mathbf{J}}\mathbf{d}\| + \|\widehat{\mathbf{J}}\mathbf{d} - \mathbf{r}\| + \|\mathbf{r} - \mathbf{J}\mathbf{d}\| \leq \epsilon\|\mathbf{d}\| + 2E\delta(p), \tag{35}$$

for $\delta(p) = 1$ if $p = 1$ and $\delta(p) = \sqrt{m}$ if $p = +\infty$, where the latter inequality comes from the fact that:

- $\|\widehat{\mathbf{J}}\mathbf{d} - \mathbf{r}\|_p \leq E$ since each solution satisfies all LP constraints,
- $\|\mathbf{r} - \mathbf{J}\mathbf{d}\|_p \leq E$ from the assumptions regarding noise,
- for any $\mathbf{x} \in \mathbb{R}^m$: $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1$ and $\|\mathbf{x}\|_2 \leq \sqrt{m}\|\mathbf{x}\|_1$,
- for any $\mathbf{x} \in \mathbb{R}^n$ and any $\mathbf{A} \in \mathbb{R}^{m \times n}$: $\|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\|_F \|\mathbf{x}\|$.

Consider the following event:

$$\mathcal{E} = \{\exists \widehat{\mathbf{J}} \in \mathbb{R}^{m \times n} : \|\widehat{\mathbf{J}} - \mathbf{J}\|_F > 2D(n) \wedge \mathcal{A}(\widehat{\mathbf{J}})\}, \tag{36}$$

where $\mathcal{A}(\widehat{\mathbf{J}})$ is the event that $\widehat{\mathbf{J}}$ is a feasible solution of the LP. Note that we have:

$$\mathcal{E} \subseteq \{\exists \widehat{\mathbf{J}} \in \mathbb{R}^{m \times n} : \|\widehat{\mathbf{J}} - \mathbf{J}\|_F > 2D(n) \wedge \mathcal{C}(\widehat{\mathbf{J}})\}, \tag{37}$$

where

$$\mathcal{C}(\widehat{\mathbf{J}})) = \{\forall_{\mathbf{d}}[\|\mathbf{d}\| > \frac{1}{\epsilon} \vee (\|\mathbf{d}\| \leq \frac{1}{\epsilon}) \wedge \|\widehat{\mathbf{J}}_{\mathrm{round}}\mathbf{d} - \mathbf{J}\mathbf{d}\| \leq \epsilon\|\mathbf{d}\| + 2E\delta(p)]\} \tag{38}$$

Notice that since $\|\widehat{\mathbf{J}}_{\mathrm{round}} - \widehat{\mathbf{J}}\|_F \leq \epsilon$ and $\epsilon < D(n)$, we get after substitution: $\eta(n) = E\delta(p)$

$$\mathcal{E} \subseteq \{\exists \widehat{\mathbf{J}}_{\mathrm{round}} \in \mathbb{R}^{m \times n} : \|\widehat{\mathbf{J}}_{\mathrm{round}} - \mathbf{J}\|_F > D(n) \wedge \mathcal{F}(\widehat{\mathbf{J}}_{\mathrm{round}})\}, \tag{39}$$

where

$$\mathcal{F}(\widehat{\mathbf{J}}_{\mathrm{round}}) = \{\forall_{\mathbf{d}}[\|\mathbf{d}\| > \frac{1}{\epsilon} \vee (\|\mathbf{d}\| \leq \frac{1}{\epsilon}) \wedge \|\widehat{\mathbf{J}}_{\mathrm{round}}\mathbf{d} - \mathbf{J}\mathbf{d}\| \leq \epsilon\|\mathbf{d}\| + 2E\delta(p)]\}. \tag{40}$$

Thus, taking the union bound over all N matrices $\widehat{\mathbf{J}}_{\mathrm{round}}$, we conclude that

$$\mathbb{P}[\mathcal{E}] \leq N((\kappa + 1 - \delta)^k). \tag{41}$$

Thus with probability at least $1 - N((\kappa + 1 - \delta)^k)$ every solution of the LP is within Frobenius norm $2D(n)$ from the exact Jacobian $\mathbf{J}$. Note that $\kappa = o_n(1)$, thus there exists some constant $\rho > 0$ such that with probability at least $1 - N(1 - \rho)^k$ every solution of the LP is within Frobenius norm distance $2D(n)$ from the exact Jacobian $\mathbf{J}$. Thus one can take $k = C\log(N)$ measurement vectors and that completes the proof. □

Proof of Lemma 3.1

Proof. The proof is based on the Caro-Wei probabilistic analysis. Order the vertices of the graph randomly and apply $\mathrm{GreedyColoring}$ algorithm. Take the first stable set from the list I and call its size X. Notice that $\mathbb{E}[X] = \sum_{v \in V(G)} \frac{1}{1+deg(v)}$. Indeed, X can be written as the sum of indicator random variables $I[v]$ for $v \in V(G)$ and where $I[v] = 0$ if v is in the first stable set and is 0 otherwise. Thus $\mathbb{E}[I[v]] = p(v)$, where $p(v)$ is the probability that v is in the first stable set. Notice that this probability is exactly $p = \sum_{v \in V(G)} \frac{1}{1+deg(v)}$, since v is in the first stable set only if it is the first vertex in the induced ordering of the set consisting of v and vertices adjacent to it. This ordering is still random and thus the result follows. Thus there exists a stable set of size $\mathbb{E}$. This set can be excluded from $V(G)$ and the procedure can be repeated. That immediately leads to the valid coloring of the vertices of the graph (each stable set is colored with a different color) and the corresponding chromatic property. Note that the procedure for finding all these stable sets is what $\mathrm{GreedyColoring}$ algorithm is doing. The only difference is that the algorithm computes sets of good sizes just on expectation. To boost the probability of success, the algorithm can be repeated some number of times even though, as previously reported, in practice this is not required. □