A. Proof of Lemma 1

By the definition of the Lovász extension, for each $i \in [r]$, we have

$$f_i(x) = \max_{y^{(i)} \in B(F_i)} \langle y^{(i)}, x \rangle.$$

Therefore

$$\begin{aligned}
&\min_{x \in \mathbb{R}^n} \sum_{i=1}^{r} \left(f_i(x) + \frac{1}{2r} \|x\|^2 \right) \\
&= \min_{x \in \mathbb{R}^n} \sum_{i=1}^{r} \left(\max_{y^{(i)} \in B(F_i)} \langle y^{(i)}, x \rangle + \frac{1}{2r} \|x\|^2 \right) \\
&= \min_{x \in \mathbb{R}^n} \max_{y^{(1)} \in B(F_1), \dots, y^{(r)} \in B(F_r)} \sum_{i=1}^{r} \left(\langle y^{(i)}, x \rangle + \frac{1}{2r} \|x\|^2 \right) \\
&= \max_{y^{(1)} \in B(F_1), \dots, y^{(r)} \in B(F_r)} \min_{x \in \mathbb{R}^n} \sum_{i=1}^{r} \left(\langle y^{(i)}, x \rangle + \frac{1}{2r} \|x\|^2 \right) \\
&= \max_{y^{(1)} \in B(F_1), \dots, y^{(r)} \in B(F_r)} -\frac{1}{2} \left\| \sum_{i=1}^{r} y^{(i)} \right\|^2 .
\end{aligned}$$

On the third line, we have used the fact that the function $\langle y, x \rangle + (1/2r) \|x\|^2$ is convex in x and linear in y, which allows us to exchange the min and the max (see for example Corollary 37.3.2 in Rockafellar (Rockafellar, 1970)). On the fourth line, we have used the fact that the minimum is achieved at $x = -\sum_{i=1}^{r} y^{(i)}$.

B. Proofs omitted from Section 2

If $x \in \mathbb{R}^{nr}$ and $\mathcal{X}$ is a subspace of $\mathbb{R}^{nr}$, we let $\Pi_{\mathcal{X}}(x)$ denote the projection of x on $\mathcal{X}$, that is, $\Pi_{\mathcal{X}}(x) = \arg\min_{z \in \mathbb{R}^{nr}} \|x - z\|$. We let $\mathcal{X}^{\perp}$ denote the orthogonal complement of the subspace $\mathcal{X}$.

Proposition 9. *For any point $x \in \mathbb{R}^{nr}$, $\Pi_{\mathcal{Q}^{\perp}}(x) = S^T S x$ and thus $\Pi_{\mathcal{Q}}(x) = x - S^T S x$.*

Proof: Since $\mathcal{Q}$ is the null space of S, $\mathcal{Q}^{\perp}$ is the row space of S. Since the rows of S are orthonormal, they form a basis for $\mathcal{Q}^{\perp}$. Therefore, if we let $v_1, \dots, v_n$ denote the rows of S, we have

$$\Pi_{\mathcal{Q}^{\perp}}(x) = \sum_{i=1}^{n} \langle x, v_i \rangle v_i = S^T S x.$$

□

Proposition 10. *The set of all optimal solutions to (Prox-DSM) is equal to E.*

Proof: We have

$$\begin{aligned}
d(\mathcal{P}, \mathcal{Q}) &= \min_{y \in \mathcal{P}} \|y - \Pi_{\mathcal{Q}}(y)\| \\
&= \min_{y \in \mathcal{P}} \left\| S^T S y \right\| \qquad \langle\langle \text{By Proposition 9} \rangle\rangle \\
&= \min_{y \in \mathcal{P}} \|Sy\| .
\end{aligned}$$

Since (Prox-DSM) is the problem $\min_{y \in \mathcal{P}} r \|Sy\|^2$, E is the set of all optimal solutions to (Prox-DSM). □

Proposition 11. *Let $y \in \mathbb{R}^{nr}$ and let $p \in E$. We have $d(y, \mathcal{Q}') = \|S(y - p)\|$.*

Proof: Since $\mathcal{Q}' = \mathcal{Q} - v$, we have

$$\begin{aligned}
d(y, \mathcal{Q}') &= d(y + v, \mathcal{Q}) \\
&= \|\Pi_{\mathcal{Q}^{\perp}}(y + v)\| \\
&= \left\| S^T S(y + v) \right\| \qquad \langle\langle \text{By Proposition 9} \rangle\rangle \\
&= \left\| S^T S(y - S^T S p) \right\| \qquad \langle\langle \text{Since } v = -S^T S p \rangle\rangle \\
&= \left\| S^T S(y - p) \right\| \qquad \langle\langle \text{Since } S S^T = I_n \rangle\rangle \\
&= \|S(y - p)\| .
\end{aligned}$$

□

Proof of Equation (3): We have

$$\begin{aligned}
&g(y_{k+1}) \\
=& g(y_k) + \int_0^1 \langle y_{k+1} - y_k, \nabla g(y_k + t(y_{k+1} - y_k)) \rangle dt \\
=& g(y_k) + \langle \nabla g(y_k), y_{k+1} - y_k \rangle + \int_0^1 \big\langle y_{k+1} - y_k, \\
& \quad \nabla g(y_k + t(y_{k+1} - y_k)) - \nabla g(y_k) \big\rangle dt \\
=& g(y_k) + \left\langle \nabla_{i_k} g(y_k), y_{k+1}^{(i_k)} - y_k^{(i_k)} \right\rangle + \int_0^1 \big\langle y_{k+1}^{(i_k)} - y_k^{(i_k)}, \\
& \nabla_{i_k} g(y_k + t(y_{k+1} - y_k)) - \nabla_{i_k} g(y_k) \big\rangle dt \\
\leq& g(y_k) + \left\langle \nabla_{i_k} g(y_k), y_{k+1}^{(i_k)} - y_k^{(i_k)} \right\rangle + \int_0^1 \left\| y_{k+1}^{(i_k)} - y_k^{(i_k)} \right\| \\
& \|\nabla_{i_k} g(y_k + t(y_{k+1} - y_k)) - \nabla_{i_k} g(y_k)\| \, dt \\
\leq& g(y_k) + \left\langle \nabla_{i_k} g(y_k), y_{k+1}^{(i_k)} - y_k^{(i_k)} \right\rangle + \\
& \int_0^1 L_{i_k} \left\| y_{k+1}^{(i_k)} - y_k^{(i_k)} \right\|^2 t dt \\
=& g(y_k) + \left\langle \nabla_{i_k} g(y_k), y_{k+1}^{(i_k)} - y_k^{(i_k)} \right\rangle + \frac{L_{i_k}}{2} \left\| y_{k+1}^{(i_k)} - y_k^{(i_k)} \right\|^2 .
\end{aligned}$$

On the third line, we have used the fact that y_k and y_{k+1} agree on all coordinate blocks except the i_k-th block. On the fourth line, we have used the Cauchy-Schwartz inequality. On the fifth line, we have used the fact that $\nabla_{i_k} g(\cdot)$ is L_{i_k}-Lipschitz. □

Proof of Equation (4): We have

$$\begin{aligned}
&\|y_{k+1} - y^*\|^2 \\
&= \|y_k - y^*\|^2 + \|y_{k+1} - y_k\|^2 + 2\langle y_k - y^*, y_{k+1} - y_k \rangle \\
&= \|y_k - y^*\|^2 - \|y_{k+1} - y_k\|^2 + 2 \langle y_{k+1} - y^*, y_{k+1} - y_k \rangle
\end{aligned}$$

$$
\begin{aligned}
&= \|y_k - y^*\|^2 - \left\|y_{k+1}^{(i_k)} - y_k^{(i_k)}\right\|^2 + \\
&\quad 2\left\langle y_{k+1}^{(i_k)} - (y^*)^{(i_k)}, y_{k+1}^{(i_k)} - y_k^{(i_k)}\right\rangle \\
&\overset{(2)}{\leq} \|y_k - y^*\|^2 - \left\|y_{k+1}^{(i_k)} - y_k^{(i_k)}\right\|^2 + \\
&\quad \frac{2}{L_{i_k}}\left\langle \nabla_{i_k} g(y_k), (y^*)^{(i_k)} - y_{k+1}^{(i_k)}\right\rangle \\
&= \|y_k - y^*\|^2 + \frac{2}{L_{i_k}}\left\langle \nabla_{i_k} g(y_k), (y^*)^{(i_k)} - y_k^{(i_k)}\right\rangle \\
&- \frac{2}{L_{i_k}}\left(\frac{L_{i_k}}{2}\left\|y_{k+1}^{(i_k)} - y_k^{(i_k)}\right\|^2 + \left\langle \nabla_{i_k} g(y_k), y_{k+1}^{(i_k)} - y_k^{(i_k)}\right\rangle\right) \\
&\overset{(3)}{\leq} \|y_k - y^*\|^2 + \frac{2}{L_{i_k}}\left\langle \nabla_{i_k} g(y_k), (y^*)^{(i_k)} - y_k^{(i_k)}\right\rangle - \\
&\quad \frac{2}{L_{i_k}}\left(g(y_{k+1}) - g(y_k)\right). \qquad (10)
\end{aligned}
$$

On the third line, we have used the fact that y_k and y_{k+1} agree on all coordinate blocks except the i_k-th block. On the fourth line, we have used the inequality (2) with $z = (y^*)^{(i_k)}$. On the last line, we have used inequality (3). $\square$