Genetics Research

A unified approach to characterize and conserve adaptive and neutral genetic diversity in subdivided populations.

Robin Wellmann*, Jörn Bennewitz, Theo Meuwissen

May 13, 2014

Appendix: Definition of the core set

Assume that $N-1$ individuals of the core set have already been created and let b be the breed of the Nth individual. For creating the Nth individual of this idealized offspring population, two gametes are randomly chosen from all individuals of breed b from the current generation. This procedure defines a sequence $\mathcal{C}=(\mathcal{C}_N)_{N\in\mathbb{N}}$ of offspring populations, whereby C_N consists of the first N individuals from the core set. Because of random mendelian sampling, $\mathcal{C}_N$ is a random offspring population. Thus, for a function D measuring some property of a population, the value $D(\mathcal{C}_N)$ is also random, but it may convergue almost surely for $N\to\infty$. In this case, we are interested in the limit $D(\mathcal{C})=\lim_{N\to\infty}D(\mathcal{C}_N)$ and $D(\mathcal{C})$ is said to be the value of D for core set $\mathcal{C}$. The value of D for a set $\mathcal{S}$ of breeds is defined as the maximum value D can achieve in a core set if only breeds from $\mathcal{S}$ are allowed to have nonzero contributions. That is,

$$D(\mathcal{S})=\sup\{D(\mathcal{C}):\mathcal{C}\text{ is a core set with }c_b=0\text{ for }b\notin\mathcal{S}\}$$

Appendix: Proofs

Equation 1:

$$TTD_t(\mathcal{C}) \quad = \quad \mathbf{c}^T\mathbf{V}_{\mathbf{A}t}+\mathbf{c}^T\left(\frac{1}{2}(\overline{\mathbf{g}}_t^2\mathbf{1}^T-2\overline{\mathbf{g}}_t\overline{\mathbf{g}}_t^T+\mathbf{1}\overline{\mathbf{g}}_t^{2T})\right)\mathbf{c}. \qquad (1)$$

Proof:

*corresponding author: University of Hohenheim, Department of Animal Husbandry and Animal Breeding, D-70599 Stuttgart, Germany. Email: r.wellmann@uni-hohenheim.de

We have

$$
\begin{aligned}
TTD_t(\mathcal{C}_N) &= \frac{1}{N}\sum_{j\in\mathcal{C}_N}(g_{tj}-\mu_{g_t})^2 \\
&= \frac{1}{N}\sum_{j\in\mathcal{C}_N} g_{tj}^2-\mu_{g_t}^2 \\
&= \frac{1}{N}\mathbf{1}^T\mathrm{diag}(\mathbf{g}_t\mathbf{g}_t^T)-(\frac{1}{N}\mathbf{1}^T\mathbf{g}_t)^2 \\
&= \frac{1}{N}\mathbf{1}^T\mathrm{diag}(\mathbf{g}_t\mathbf{g}_t^T)-\frac{1}{N^2}\mathbf{1}^T\mathbf{g}_t\mathbf{g}_t^T\mathbf{1},
\end{aligned}
$$

where $\mathbf{g}_t \in I\!R^N$ contains the genotypic values of all individuals from $\mathcal{C}_N$ for trait t.

Let $B_b \subset \{1,...,N\}$ be the set of individuals in the offspring population $\mathcal{C}_N$ belonging to breed b, N_b is the number of individuals from breed b in the offspring population $\mathcal{C}_N$, and $\mathbf{1}_b \in I\!R^N$ is a vector with zeros and ones, where $1_{bi} = 1$ if individual i from the offspring population belongs to breed b. Let $\tilde{\mathbf{S}}_o = (\frac{1}{N_1}\mathbf{1}_1,...,\frac{1}{N_B}\mathbf{1}_B) \in I\!R^{N\times B}$ and $c_b = \frac{N_b}{N}$. Since

$$
\frac{1}{N}\mathbf{1} = \tilde{\mathbf{S}}_o\mathbf{c} \in I\!R^N,
$$

we have

$$
TTD_t(\mathcal{C}_N) = \mathbf{c}^T\mathbf{m}_{Nt}-\mathbf{c}^T\mathbf{M}_{Nt}\mathbf{c},
$$

with

$$
\begin{aligned}
\mathbf{M}_{Nt} &= \tilde{\mathbf{S}}_o^T\mathbf{g}_t\mathbf{g}_t^T\tilde{\mathbf{S}}_o \\
\mathbf{m}_{Nt} &= \tilde{\mathbf{S}}_o^T\mathrm{diag}(\mathbf{g}_t\mathbf{g}_t^T).
\end{aligned}
$$

We have

$$
\begin{aligned}
M_{Ntbl} &= \frac{1}{N_bN_l}\mathbf{1}_b^T\mathbf{g}_t\mathbf{g}_t^T\mathbf{1}_l = \left(\frac{1}{N_b}\sum_{i\in B_b} g_{ti}\right)\left(\frac{1}{N_l}\sum_{j\in B_l} g_{tj}\right), \\
m_{Ntb} &= \frac{1}{N_b}\mathbf{1}_b^T\mathrm{diag}(\mathbf{g}_t\mathbf{g}_t^T) = \frac{1}{N_b}\sum_{i\in B_b} g_{ti}^2.
\end{aligned}
$$

The definition of TTD shows that adding a constant to all genotypic values does not change the value of the objective function, so the vector with genotypic values is

$$
\mathbf{g}_t = (\mathbf{Z}-2\mathbf{1}\mathbf{p}_0^T)\mathbf{a}_t,
$$

where $\mathbf{a}_t \in I\!R^M$ is the vector with true SNP effects ($a_{tm} = 0$ if SNP m is not a QTL), and $\mathbf{p}_0 \in I\!R^M$ is a vector containing arbitrary values. The matrix $\mathbf{Z} \in I\!R^{N\times M}$ is the gene content matrix for the 1-alleles with entries 0,1, and 2. We can write $\mathbf{Z}_i = \mathbf{m}_i + \mathbf{s}_i$, where $\mathbf{Z}_i^T$ is the ith row of matrix $\mathbf{Z}$, $\mathbf{m}_i \in I\!R^M$ is the vector with maternal SNP alleles, $\mathbf{s}_i \in I\!R^M$ is the vector with paternal SNP alleles of individual i. We have

$$
\begin{aligned}
m_{Ntb} &= \frac{1}{N_b}\sum_{i\in B_b} g_{ti}^2 \\
&= \frac{1}{N_b}\sum_{i\in B_b} \mathbf{a}_t^T(\mathbf{Z}_i - 2\mathbf{p}_0)(\mathbf{Z}_i^T - 2\mathbf{p}_0^T)\mathbf{a}_t \\
&= \mathbf{a}_t^T\left(\frac{1}{N_b}\sum_{i\in B_b}(\mathbf{Z}_i - 2\mathbf{p}_0)(\mathbf{Z}_i^T - 2\mathbf{p}_0^T)\right)\mathbf{a}_t \\
&= \mathbf{a}_t^T\left(\frac{1}{N_b}\sum_{i\in B_b}(\mathbf{m}_i + \mathbf{s}_i - 2\mathbf{p}_0)(\mathbf{m}_i^T + \mathbf{s}_i^T - 2\mathbf{p}_0^T)\right)\mathbf{a}_t \\
&= \mathbf{a}_t^T\left(\frac{1}{N_b}\sum_{i\in B_b}(\mathbf{m}_i + \mathbf{s}_i)(\mathbf{m}_i^T + \mathbf{s}_i^T)\right)\mathbf{a}_t \\
&+ \mathbf{a}_t^T\left(\frac{1}{N_b}\sum_{i\in B_b}4\mathbf{p}_0\mathbf{p}_0^T - (\mathbf{m}_i + \mathbf{s}_i)2\mathbf{p}_0^T - 2\mathbf{p}_0(\mathbf{m}_i + \mathbf{s}_i)^T\right)\mathbf{a}_t \\
&= \mathbf{a}_t^T\left(\frac{1}{N_b}\sum_{i\in B_b}\mathbf{m}_i\mathbf{m}_i^T + \mathbf{m}_i\mathbf{s}_i^T + \mathbf{s}_i\mathbf{m}_i^T + \mathbf{s}_i\mathbf{s}_i^T\right)\mathbf{a}_t \\
&+ \mathbf{a}_t^T\left(\frac{1}{N_b}\sum_{i\in B_b}4\mathbf{p}_0\mathbf{p}_0^T - 2\mathbf{m}_i\mathbf{p}_0^T - 2\mathbf{s}_i\mathbf{p}_0^T - 2\mathbf{p}_0\mathbf{m}_i^T - 2\mathbf{p}_0\mathbf{s}_i^T\right)\mathbf{a}_t.
\end{aligned}
$$

Note that

$$\lim_{N_b\to\infty}\frac{1}{N_b}\sum_{i\in B_b}\mathbf{m}_i\mathbf{p}_0^T = \lim_{N_b\to\infty}\frac{1}{N_b}\sum_{i\in B_b}\mathbf{s}_i\mathbf{p}_0^T = \mathbf{p}_b\mathbf{p}_0^T.$$

Since the offspring population was created by random mating within populations, the maternal and the paternal alleles of an individual were independently chosen from the current population, so

$$\lim_{N_b\to\infty}\frac{1}{N_b}\sum_{i\in B_b}\mathbf{m}_i\mathbf{s}_i^T = \lim_{N_b\to\infty}\frac{1}{N_b}\sum_{i\in B_b}\mathbf{s}_i\mathbf{m}_i^T = \mathbf{p}_b\mathbf{p}_b^T.$$

Let $\mathbf{H}_b$ be random M-vector containing the SNP alleles of a gamete randomly chosen from individuals of breed b in the current population. Since maternal and paternal alleles are identically distributed, we have

$$\lim_{N_b\to\infty}\frac{1}{N_b}\sum_{i\in B_b}\mathbf{m}_i\mathbf{m}_i^T = \lim_{N_b\to\infty}\frac{1}{N_b}\sum_{i\in B_b}\mathbf{s}_i\mathbf{s}_i^T = E(\mathbf{H}_b\mathbf{H}_b^T),$$

and $E(\mathbf{H}_b) = \mathbf{p}_b$. Since the additive variance of trait t in population b is $V_{Atb} =$

$2\mathbf{a}_t^T cov(\mathbf{H}_b)\mathbf{a}_t$ and $\overline{g}_{tb} = (2\mathbf{p}_b - 2\mathbf{p}_0)^T\mathbf{a}_t$, it follows that

$$\begin{aligned}
m_{tb} &= \lim_{N\to\infty} m_{Ntb} \\
&= \mathbf{a}_t^T\left(E(\mathbf{H}_b\mathbf{H}_b^T) + \mathbf{p}_b\mathbf{p}_b^T + \mathbf{p}_b\mathbf{p}_b^T + E(\mathbf{H}_b\mathbf{H}_b^T)\right)\mathbf{a}_t \\
&+ \mathbf{a}_t^T\left(4\mathbf{p}_0\mathbf{p}_0^T - 2\mathbf{p}_b\mathbf{p}_0^T - 2\mathbf{p}_b\mathbf{p}_0^T - 2\mathbf{p}_0\mathbf{p}_b^T - 2\mathbf{p}_0\mathbf{p}_b^T\right)\mathbf{a}_t \\
&= \mathbf{a}_t^T\left(2E(\mathbf{H}_b\mathbf{H}_b^T) + 2\mathbf{p}_b\mathbf{p}_b^T + 4\mathbf{p}_0\mathbf{p}_0^T - 4\mathbf{p}_b\mathbf{p}_0^T - 4\,\mathbf{p}_0\mathbf{p}_b^T\right)\mathbf{a}_t \\
&= \mathbf{a}_t^T\left(2cov(\mathbf{H}_b) + 4\mathbf{p}_b\mathbf{p}_b^T + 4\mathbf{p}_0\mathbf{p}_0^T - 4\mathbf{p}_b\mathbf{p}_0^T - 4\,\mathbf{p}_0\mathbf{p}_b^T\right)\mathbf{a}_t \\
&= 2\mathbf{a}_t^T cov(\mathbf{H}_b)\mathbf{a}_t + \mathbf{a}_t^T(2\mathbf{p}_b - 2\mathbf{p}_0)(2\mathbf{p}_b - 2\mathbf{p}_0)^T\mathbf{a}_t \\
&= V_{Atb} + \overline{g}_{tb}^2.
\end{aligned}$$

Moreover,

$$\begin{aligned}
M_{tbl} &= \lim_{N_b,N_l\to\infty} M_{Ntbl} \\
&= \lim_{N_b,N_l\to\infty}\left(\frac{1}{N_b}\sum_{i\in B_b} g_{ti}\right)\left(\frac{1}{N_l}\sum_{j\in B_l} g_{tj}\right) \\
&= \overline{g}_{tb}\overline{g}_{tl}.
\end{aligned}$$

Thus,

$$TTD_t(\mathcal{C}) = \mathbf{c}^T\mathbf{m}_t - \mathbf{c}^T\mathbf{M}_t\mathbf{c},$$

where

$$\begin{aligned}
\mathbf{M}_t &= \overline{\mathbf{g}}_t\overline{\mathbf{g}}_t^T, \\
m_{tb} &= V_{Atb} + \overline{g}_{tb}^2, \\
\overline{g}_{tb} &= (2\mathbf{p}_b - 2\mathbf{p}_0)^T\mathbf{a}_t.
\end{aligned}$$

Since $\mathbf{1}^T\mathbf{c} = 1$, it follows that

$$\begin{aligned}
TTD_t(\mathcal{C}) &= \mathbf{c}^T(\mathbf{V_A}_t + \overline{\mathbf{g}}_t^2) - \mathbf{c}^T\overline{\mathbf{g}}_t\overline{\mathbf{g}}_t^T\mathbf{c}, \\
&= \mathbf{c}^T\mathbf{V_A}_t + \frac{1}{2}\left(\mathbf{c}^T\overline{\mathbf{g}}_t^2\mathbf{1}^T\mathbf{c} + \mathbf{c}^T\mathbf{1}\overline{\mathbf{g}}_t^{2T}\mathbf{c} - 2\mathbf{c}^T\overline{\mathbf{g}}_t\overline{\mathbf{g}}_t^T\mathbf{c}\right) \\
&= \mathbf{c}^T\mathbf{V_A}_t + \mathbf{c}^T\left(\frac{1}{2}\left(\overline{\mathbf{g}}_t^2\mathbf{1}^T - 2\overline{\mathbf{g}}_t\overline{\mathbf{g}}_t^T + \mathbf{1}\overline{\mathbf{g}}_t^{2T}\right)\right)\mathbf{c},
\end{aligned}$$

Equation 2:

$$\begin{aligned}
NTD_t(\mathcal{C}) &= \mathbf{c}^T(V_t(\mathbf{1} + \mathbf{F})) - \mathbf{c}^T(2V_t\mathbf{f})\mathbf{c} \qquad (2) \\
&= V_t\mathbf{c}^T(\mathbf{1} - \mathbf{F}) + V_t\mathbf{c}^T\left(\mathbf{F}\mathbf{1}^T - 2\mathbf{f} + \mathbf{1}\mathbf{F}^T\right)\mathbf{c},
\end{aligned}$$

Proof:

From Equation (1) it follows that

$$NTD_t(\mathcal{C}) = \mathbf{c}^T E(\mathbf{V_A}_t) + \mathbf{c}^T \left(\frac{1}{2} E(\overline{\mathbf{g}}_t^2 \mathbf{1}^T - 2\overline{\mathbf{g}}_t \overline{\mathbf{g}}_t^T + \mathbf{1}\overline{\mathbf{g}}_t^{2T})\right) \mathbf{c}.$$

From conditions A) and B) we obtain

$$\begin{aligned} (1 - \mathbf{F})V_t &= E(\mathbf{V_A}_t), \\ \frac{V_t}{4\alpha}\left(\mathbf{F}\mathbf{1}^T - 2\mathbf{f} + \mathbf{1}\mathbf{F}^T\right) &= \frac{1}{2} E\left(\overline{\mathbf{g}}_t^2 \mathbf{1}^T - 2\overline{\mathbf{g}}_t \overline{\mathbf{g}}_t^T + \mathbf{1}\overline{\mathbf{g}}_t^{2T}\right). \end{aligned}$$

Thus,

$$NTD_t(\mathcal{C}) = V_t \mathbf{c}^T (1 - \mathbf{F}) + \frac{V_t}{4\alpha} \mathbf{c}^T \left(\mathbf{F}\mathbf{1}^T - 2\mathbf{f} + \mathbf{1}\mathbf{F}^T\right) \mathbf{c}.$$

The analogous equation obtained by Bennewitz and Meuwissen (2005b) using a pedigree based approach can be written as

$$NTD_t^{Ped}(\mathcal{C}) = V_t \mathbf{c}^T (\mathbf{1} - \mathbf{F_P}) + V_t \mathbf{c}^T \left(\mathbf{F_P}\mathbf{1}^T - 2\mathbf{f_P} + \mathbf{1}\mathbf{F_P}^T\right) \mathbf{c},$$

where $\mathbf{f_P}$ denotes a pedigree based kinship matrix, $\mathbf{F_P} = diag(\mathbf{f_P})$, and V_t is a scaling parameter. Since we would like that the marker based kinship matrix has similar properties as the pedigree based kinship matrix, we use $\alpha = \frac{1}{4}$.

In the following we derive the explicit formulas for computing $\mathbf{f}$. From condition A) we get

$$\begin{aligned} f_{bb} &= 1 - \frac{E(V_{Atb})}{V_t} \\ &= 1 - \frac{\sum_{m=1}^{M} 2p_{bm}(1 - p_{bm})E(a_{tm}^2)}{V_t} \\ &= 1 - \frac{p_{QTL}\sigma_{a_t}^2}{V_t} \sum_{m=1}^{M} 2p_{bm}(1 - p_{bm}) \\ &= 1 - \frac{\tilde{V}_t}{V_t} \frac{1}{M} 2\mathbf{p}_b^T(\mathbf{1} - \mathbf{p}_b), \end{aligned}$$

where $\tilde{V}_t = p_{QTL}\sigma^2_{a_t}M$. From condition B) we get for $\alpha = \frac{1}{4}$:

$$\begin{aligned}
f_{bl} &= \frac{f_{bb}+f_{ll}}{2} - \frac{1}{4V_t}E(\overline{\mathbf{g}}^2_{tb} - 2\overline{\mathbf{g}}_{tb}\overline{\mathbf{g}}_{tl} + \overline{\mathbf{g}}^2_{tl}) \\
&= \frac{f_{bb}+f_{ll}}{2} - \frac{E((\overline{\mathbf{g}}_{tb} - \overline{\mathbf{g}}_{tl})^2)}{4V_t} \\
&= \frac{f_{bb}+f_{ll}}{2} - \frac{E(((2\mathbf{p}_b - 2\mathbf{p}_0)^T\mathbf{a}_t - (2\mathbf{p}_l - 2\mathbf{p}_0)^T\mathbf{a}_t)^2)}{4V_t} \\
&= \frac{f_{bb}+f_{ll}}{2} - \frac{E(((2\mathbf{p}_b - 2\mathbf{p}_l)^T\mathbf{a}_t)^2)}{4V_t} \\
&= \frac{f_{bb}+f_{ll}}{2} - \frac{E((2\mathbf{p}_b - 2\mathbf{p}_l)^T\mathbf{a}_t\mathbf{a}_t^T(2\mathbf{p}_b - 2\mathbf{p}_l))}{4V_t} \\
&= \frac{f_{bb}+f_{ll}}{2} - \frac{(2\mathbf{p}_b - 2\mathbf{p}_l)^T Cov(\mathbf{a}_t)(2\mathbf{p}_b - 2\mathbf{p}_l)}{4V_t} \\
&= \frac{f_{bb}+f_{ll}}{2} - \frac{p_{QTL}\sigma^2_{a_t}M}{4V_t}\frac{1}{M}(2\mathbf{p}_b - 2\mathbf{p}_l)^T(2\mathbf{p}_b - 2\mathbf{p}_l) \\
&= \frac{f_{bb}+f_{ll}}{2} - \frac{\tilde{V}_t}{V_t}\frac{1}{M}(\mathbf{p}_b - \mathbf{p}_l)^T(\mathbf{p}_b - \mathbf{p}_l)
\end{aligned}$$

Thus,

$$\begin{aligned}
f_{bl} &= \frac{1 - \frac{\tilde{V}_t}{V_t}\frac{1}{M}2\mathbf{p}_b^T(\mathbf{1} - \mathbf{p}_b) + 1 - \frac{\tilde{V}_t}{V_t}\frac{1}{M}2\mathbf{p}_l^T(\mathbf{1} - \mathbf{p}_l)}{2} - \frac{\tilde{V}_t}{V_t}\frac{1}{M}(\mathbf{p}_b - \mathbf{p}_l)^T(\mathbf{p}_b - \mathbf{p}_l) \\
&= 1 - \frac{\tilde{V}_t}{V_t}\frac{1}{M}\left(\mathbf{p}_b^T(\mathbf{1} - \mathbf{p}_b) + \mathbf{p}_l^T(\mathbf{1} - \mathbf{p}_l) + (\mathbf{p}_b - \mathbf{p}_l)^T(\mathbf{p}_b - \mathbf{p}_l)\right) \\
&= 1 - \frac{\tilde{V}_t}{V_t}\frac{1}{M}\left(\mathbf{p}_b^T\mathbf{1} - \mathbf{p}_b^T\mathbf{p}_b + \mathbf{p}_l^T\mathbf{1} - \mathbf{p}_l^T\mathbf{p}_l + \mathbf{p}_b^T\mathbf{p}_b - 2\mathbf{p}_b^T\mathbf{p}_l + \mathbf{p}_l^T\mathbf{p}_l\right) \\
&= 1 - \frac{\tilde{V}_t}{V_t}\frac{1}{M}\left(\mathbf{p}_b^T\mathbf{1} + \mathbf{p}_l^T\mathbf{1} - 2\mathbf{p}_b^T\mathbf{p}_l\right) \\
&= 1 - \frac{\tilde{V}_t}{2V_t}\frac{1}{M}\left(\mathbf{1}^T\mathbf{1} - (2\mathbf{p}_b - \mathbf{1})^T(2\mathbf{p}_l - \mathbf{1})\right) \\
&= 1 - \frac{\tilde{V}_t}{2V_t}\left(1 - \frac{1}{M}(2\mathbf{p}_b - \mathbf{1})^T(2\mathbf{p}_l - \mathbf{1})\right)
\end{aligned}$$

The scale parameter V_t may be chosen arbitrarily. However, in order to ensure that $f_{bb} \geq 0$ for every vector $\mathbf{p}_b$ containing allele frequencies, $V_t \geq \frac{\tilde{V}_t}{2}$ should be chosen. In the paper we used

$$V_t = \frac{\tilde{V}_t}{\kappa} = \frac{p_{QTL}\sigma^2_{a_t}M}{\kappa}$$

with $\kappa = 2$ in order to get a high variability of the marker based kinships. In this case, the formula for f_{bl} can be further simplified:

$$f_{bl} = \frac{1}{M}(2\mathbf{p}_b - \mathbf{1})^T(2\mathbf{p}_l - \mathbf{1}).$$

Thus,

$$\mathbf{f} = \frac{1}{M}\sum_{m=1}^{M}(2\mathbf{p}_{(m)} - \mathbf{1})(2\mathbf{p}_{(m)} - \mathbf{1})^T.$$

Equation 4:

$$NGD(\mathcal{C}) = \frac{1}{M}\sum_{m=1}^{M} 2\,\mathbf{c}^T\mathbf{p}_{(m)}\,(1 - \mathbf{c}^T\mathbf{p}_{(m)}) = \frac{1 - \mathbf{c}^T\mathbf{f}\mathbf{c}}{2}, \qquad (4)$$

Proof:

The equality on the right hand side holds because

$$\begin{aligned}
NGD(\mathcal{C}) &= \frac{1}{M}\sum_{m=1}^{M} 2\mathbf{c}^T\mathbf{p}_{(m)}(1 - \mathbf{c}^T\mathbf{p}_{(m)}) \\
&= \frac{1}{M}\sum_{m=1}^{M} 2\mathbf{c}^T\mathbf{p}_{(m)} - 2\mathbf{c}^T\mathbf{p}_{(m)}\mathbf{p}_{(m)}^T\mathbf{c} \\
&= \frac{1}{M}\sum_{m=1}^{M} \mathbf{c}^T\mathbf{p}_{(m)}\mathbf{1}^T\mathbf{c} + \mathbf{c}^T\mathbf{1}\mathbf{p}_{(m)}^T\mathbf{c} - 2\mathbf{c}^T\mathbf{p}_{(m)}\mathbf{p}_{(m)}^T\mathbf{c} \\
&= \frac{1}{M}\mathbf{c}^T\left(\sum_{m=1}^{M}(\mathbf{p}_{(m)}\mathbf{1}^T + \mathbf{1}\mathbf{p}_{(m)}^T - 2\mathbf{p}_{(m)}\mathbf{p}_{(m)}^T)\right)\mathbf{c} \\
&= \frac{1}{M}\mathbf{c}^T\left(\sum_{m=1}^{M}\frac{1}{2}(\mathbf{1}\mathbf{1}^T - (\mathbf{1}\mathbf{1}^T - 2\mathbf{p}_{(m)}\mathbf{1}^T - 2\mathbf{1}\mathbf{p}_{(m)}^T + 4\mathbf{p}_{(m)}\mathbf{p}_{(m)}^T))\right)\mathbf{c} \\
&= \frac{1}{M}\mathbf{c}^T\left(\sum_{m=1}^{M}\frac{1}{2}(\mathbf{1}\mathbf{1}^T - (\mathbf{1} - 2\mathbf{p}_{(m)})(\mathbf{1} - 2\mathbf{p}_{(m)})^T)\right)\mathbf{c} \\
&= \frac{1}{2}\,\mathbf{c}^T\left(\mathbf{1}\mathbf{1}^T - \frac{1}{M}\sum_{m=1}^{M}(\mathbf{1} - 2\mathbf{p}_{(m)})(\mathbf{1} - 2\mathbf{p}_{(m)})^T\right)\mathbf{c} \\
&= \frac{1}{2}\,\mathbf{c}^T\left(\mathbf{1}\mathbf{1}^T - \mathbf{f}\right)\mathbf{c} \\
&= \frac{1}{2}\left(1 - \mathbf{c}^T\mathbf{f}\mathbf{c}\right).
\end{aligned}$$