CliPP2

Yu Ding*

Department of Bioinformatics and Computational Biology, MD Anderson Cancer Center

April 22, 2025

Abstract

We develop a SCAD-penalized likelihood framework for subclone reconstruction in *multi-sample* settings ($M > 1$). We consider a set of single nucleotide variants (SNVs) across M tumor samples from a single patient. For each SNV i, we associate an unknown vector $\mathbf{p}_i \in \mathbb{R}^M$ capturing its logistic-scale parameters across all samples. A group SCAD penalty $\|\mathbf{p}_i - \mathbf{p}_j\|_2$ is then imposed to encourage clusters of SNVs with matching parameter vectors. This induces a subclonal structure across multiple samples simultaneously.

We provide an end-to-end derivation of the optimization problem, including:

- A precise formulation of the negative binomial log-likelihood for multi-sample read count data;
- A piecewise-linear (or quasi-likelihood normal) approximation that yields a quadratic surrogate for the negative log-likelihood at each iteration;
- A full ADMM (Alternating Direction Method of Multipliers) algorithm, complete with augmented Lagrangian details and closed-form (or piecewise-closed-form) updates for each subproblem;
- A group-SCAD thresholding step to update the auxiliary variables corresponding to pairwise differences in $\mathbf{p}_i$.

All notation and variables are clearly introduced and defined before use. This document focuses **exclusively** on the case where $M > 1$. For completeness, we also outline how the dimension analysis applies when many SNVs and multiple samples are present, and we discuss potential computational strategies to handle large-scale data.

Contents

*yding4@mdanderson.org

1 Introduction and Motivation

In contemporary cancer genomics, it has become increasingly common to obtain *multiple tumor samples* ($M > 1$) from the same patient. This multi-sample context can arise either from:

1. **Spatial heterogeneity**, where a surgeon or pathologist takes samples from different *regions* of the tumor. Different regions may harbor distinct subclonal populations due to local evolution.

2. **Longitudinal sampling**, where tumor biopsies are taken at different time points (e.g., before and after treatment), thus allowing tracking of emerging subclones over time.

Regardless of the precise experimental design, multi-sample data provide a richer view of *clonal evolution*. By leveraging multiple samples, one can gain insight into how a single SNV's frequency may vary across different spatial or temporal conditions, potentially identifying branching evolutionary paths or selective sweeps.

Subclone reconstruction. The core problem is to group Single Nucleotide Variants (SNVs) into clusters (subclones) that presumably correspond to distinct populations of tumor cells. Each subclone might have a characteristic *pattern of SNV frequencies (cellular prevalence)* across the M samples. If two SNVs share the same pattern across all samples, it is often hypothesized that they belong to the same subclone (or clonal lineage). Consequently, an effective computational approach attempts to *cluster* or *merge* SNVs whose multi-sample frequency vectors are identical or nearly identical.

Logistic reparametrization. When counting reads for each SNV within each sample, one typically collects read depths $\{n_{ij}\}$ and observed variant reads $\{r_{ij}\}$. A classical approach is to model r_{ij} via a binomial distribution with probability θ_{ij}. The parameter θ_{ij} is the fraction of reads (in sample j) that correspond to the variant for SNV i. A common reparametrization is to let p_{ij} be the *logistic-scale* parameter, so that

$$\theta_{ij} = \frac{e^{p_{ij}}}{1 + e^{p_{ij}}}.$$

Alternatively, with copy number/purity adjustments, one might write

$$\theta_{ij}(p_{ij}) = \frac{b_{ij}^{V}\, e^{p_{ij}}}{(1 + e^{p_{ij}})\, A_{ij}},$$

where A_{ij} is known from external copy number/purity estimates, and $\mathbf{p}_i = (p_{i1}, \ldots, p_{iM}) \in \mathbb{R}^M$.

Group SCAD penalty. A powerful method to enforce *clustering* of these vectors $\mathbf{p}_i$ across SNVs is to penalize the pairwise ℓ_2-norm difference:

$$\|\mathbf{p}_i - \mathbf{p}_j\|_2,$$

using a nonconvex penalty such as SCAD. Precisely:

$$\mathrm{SCAD}_\lambda\big(\|\mathbf{p}_i - \mathbf{p}_j\|_2\big).$$

Whenever $\mathbf{p}_i$ and $\mathbf{p}_j$ are forced to coincide, SNVs i and j share exactly the same vector across all samples, thus grouping them into one subclone.

Goal and outline of this document. This paper constructs the penalized likelihood, carefully defining every variable, and presents the derivation of an ADMM algorithm. The main challenge is that the binomial negative log-likelihood (based on θ_{ij}) is not purely quadratic, and the SCAD penalty is nonconvex. Our solution is to use a local *piecewise-linear* or *quasi-likelihood* approximation to the logistic part, ensuring a *quadratic* subproblem at each iteration. This subproblem is then tackled by ADMM with an *augmented Lagrangian*, introducing auxiliary and dual variables. Each update step turns out to have closed-form (or piecewise-closed-form) solutions, making the approach both conceptually straightforward and computationally feasible for moderate to large scale.

Below is a roadmap of the paper:

- Section 2 explicitly defines all variables and notation used in the multi-sample problem.
- Section 3 presents the negative log-likelihood for multi-sample binomial data, clarifying how θ_{ij} is related to $\mathbf{p}_i$.
- Section 4 explains the group SCAD penalty, including its piecewise definition and underlying intuition.
- Section 5 sets up the augmented Lagrangian, introduces auxiliary variables for pairwise differences $\boldsymbol{\eta}_{ij}$, and shows the ADMM splitting approach.
- Section 6 details how we locally approximate the negative log-likelihood by a *piecewise-linear* or *normal/quasi-likelihood* expansion, yielding a strictly quadratic subproblem.
- Section 7 derives each ADMM update step in detail, **including the explicit expressions of all elements of matrices B and A** within the IRLS-based approximation.
- Section 8 discusses dimension analysis, addressing how the approach scales with the number of SNVs S and samples M.
- Section 9 recapitulates the full algorithm and offers insights on implementation details such as linear algebra efficiency and convergence monitoring.
- Section 10 concludes with remarks about scalability, limitations, and future directions.

- Finally, Appendix A sketches an alternative normal/quasi-likelihood viewpoint that can also produce a local quadratic surrogate for the binomial objective.

This document is designed to be self-contained and focuses **exclusively** on the $\mathbf{M > 1}$ scenario. All symbols and variables are introduced in Section 2 before use.

2 Notation and Variables

In this section, we state every variable clearly and consistently, ensuring that no symbol is used without prior definition.

2.1 Data and Index Sets

- $\mathbf{S}$: The total number of single nucleotide variants (SNVs). We index them by $i = 1, 2, \ldots, S$.
- $\mathbf{M}$: The total number of tumor samples. We index them by $j = 1, 2, \ldots, M$. $\mathbf{M} > 1$ throughout this paper.
- r_{ij}: For SNV i in sample j, r_{ij} is the number of reads that support the variant (i.e., how many times the mutant base is observed).
- n_{ij}: For SNV i in sample j, n_{ij} is the total read count at that locus (i.e., the sum of all reads, both reference and variant).
- θ_{ij}: The *true* fraction of variant reads expected for SNV i in sample j. In practice, θ_{ij} is unknown and must be inferred.

2.2 Logistic-scale Variables

- $\mathbf{p}_i$: A vector in $\mathbb{R}^M$ that parametrizes SNV i across *all* M samples. We write
$$\mathbf{p}_i = \big(p_{i1},\, p_{i2},\, \ldots,\, p_{iM}\big).$$
- p_{ij}: The logistic-scale parameter for SNV i in sample j. We define
$$p_{ij} \in \mathbb{R},$$
which often maps to θ_{ij} via
$$\theta_{ij} = \frac{e^{p_{ij}}}{1 + e^{p_{ij}}}\,.$$

2.3 Copy Number and Purity Adjustments

Sometimes one has known or estimated copy number and purity factors, making:

$$\theta_{ij}(p_{ij}) = \frac{b_{ij}^V\, e^{p_{ij}}}{\big(1 + e^{p_{ij}}\big)\, A_{ij}},$$

where:

- b_{ij}^V: The *multiplicity* for SNV i of sample j, $b_{ij}^V = \min\Big(\Big[\frac{r_{ij}}{n_{ij}}\frac{1}{\rho_j}(\rho_j c_{ij}^T + c_{ij}^N(1-\rho_j)\Big]_0, m_{ij}^V\Big)$, where m_{ij}^V is the corresponding major copy number.
- A_{ij}: A known factor for sample j that can incorporate normal/tumor mixing and total copy number. For example, if ρ_j is tumor purity, c_{ij}^N and c_{ij}^T are normal/tumor copy numbers, then $A_{ij} = (1-\rho_j)c_{ij}^N + \rho_j\, c_{ij}^T$.

All such factors are assumed to be known or pre-estimated. In the final optimization, we solve for $\mathbf{p}_i$ given these known factors.

2.4 Negative Log-likelihood

- We assume that r_{ij} is generated via a Binomial(n_{ij}, θ_{ij}). Thus the negative log-likelihood for the entire dataset is:

$$-\ell\big(\{\mathbf{p}_i\}\big) = \sum_{i=1}^{S}\sum_{j=1}^{M}\Big[-\log\binom{n_{ij}}{r_{ij}} - r_{ij}\,\log\big(\theta_{ij}(p_{ij})\big) - (n_{ij}-r_{ij})\,\log\big(1-\theta_{ij}(p_{ij})\big)\Big].$$

  Since $-\log\binom{n_{ij}}{r_{ij}}$ does not depend on p_{ij}, it can be omitted if we only care about $\{\mathbf{p}_i\}$.

2.5 Pairwise Differences and Group SCAD

- To promote clustering among the vectors $\mathbf{p}_1, \ldots, \mathbf{p}_S$, we introduce a *group SCAD* penalty on the pairwise differences. For each (i, j) with $1 \le i < j \le S$, define

$$\|\mathbf{p}_i - \mathbf{p}_j\|_2 = \sqrt{\sum_{m=1}^{M}\big(p_{im} - p_{jm}\big)^2}.$$

- The SCAD penalty function, denoted $\mathrm{SCAD}_\lambda(\cdot)$, is a piecewise-defined function with a shape parameter $\gamma > 2$. For a scalar $t \ge 0$,

$$\mathrm{SCAD}_\lambda(t) = \begin{cases} \lambda\, t, & 0 \le t \le \lambda, \\ -\dfrac{t^2 - 2(\gamma\lambda)\,t + \lambda^2}{2(\gamma-1)}, & \lambda < t \le \gamma\,\lambda, \\ \dfrac{(\gamma+1)\,\lambda^2}{2}, & t > \gamma\,\lambda. \end{cases}$$

  We then apply this to $t = \|\mathbf{p}_i - \mathbf{p}_j\|_2$.

2.6 Overall Objective

- We form a penalized objective function:

$$\min_{\{\mathbf{p}_i\}_{i=1}^{S}} \Big\{-\ell\big(\{\mathbf{p}_i\}\big) \;+\; \sum_{1\le i<j\le S} \mathrm{SCAD}_\lambda\big(\|\mathbf{p}_i - \mathbf{p}_j\|_2\big)\Big\}.$$

- We typically solve this iteratively, approximating ℓ by a local quadratic or piecewise-linear function. We then apply ADMM to handle the nonconvex penalty on $\|\mathbf{p}_i - \mathbf{p}_j\|_2$.

3 Multi-sample Binomial Negative Log-likelihood

We now detail how the binomial negative log-likelihood arises in the multi-sample scenario. This also sets the stage for the subsequent piecewise approximation.

3.1 Model Assumption

For each SNV i and sample j, the observed number of variant-supporting reads r_{ij} is modeled as:

$$r_{ij} \;\sim\; \mathrm{Binomial}\big(n_{ij},\, \theta_{ij}\big),$$

where n_{ij} is the total read count, and

$$\theta_{ij} = \theta_{ij}(p_{ij}) = \frac{b_{ij}^{V}\, e^{p_{ij}}}{(1 + e^{p_{ij}})\, A_{ij}}$$

or $\theta_{ij} = \frac{e^{p_{ij}}}{(1+e^{p_{ij}})}$ if no copy number/purity correction is used. The essential detail is that θ_{ij} is a logistic function of p_{ij}.

3.2 Explicit Form of the Negative Log-likelihood

The binomial negative log-likelihood for each pair (i, j) (omitting constants unrelated to p_{ij}) is:

$$-\ln P(r_{ij}|p_{ij}) = -\,r_{ij}\,\ln\big(\theta_{ij}(p_{ij})\big) - (n_{ij} - r_{ij})\,\ln\big(1-\theta_{ij}(p_{ij})\big).$$

Summing over all SNVs i and all samples j:

$$-\ell\big(\{p_{ij}\}\big) = \sum_{i=1}^{S}\sum_{j=1}^{M}\Big[-\,r_{ij}\,\ln\theta_{ij}(p_{ij}) - (n_{ij} - r_{ij})\,\ln\big(1-\theta_{ij}(p_{ij})\big)\Big].$$

We will denote this collectively as

$$-\ell(\mathbf{p}),$$

where $\mathbf{p} \in \mathbb{R}^{S\,M}$ stacks all p_{ij}. Although it is convex in θ_{ij}, it is *not* globally convex in p_{ij} due to the logistic transformation. Nonetheless, it is *smooth* and thus amenable to second-order (Newton or IRLS) approximation.

3.3 Why Approximate Quadratically?

In principle, one could attempt to solve the penalized problem directly using gradient-based methods. However, the group SCAD penalty $\mathrm{SCAD}_\lambda\big(\|\mathbf{p}_i - \mathbf{p}_j\|_2\big)$ is nonconvex and introduces ℓ_2-type couplings among different $\mathbf{p}_i$. A naive gradient-based method would be challenging to converge reliably.

By constructing a *local quadratic surrogate* for $-\ell(\mathbf{p})$ around the current iterate, we gain the ability to separate the problem into simpler updates. This approach is reminiscent of *Iteratively Reweighted Least Squares (IRLS)* in logistic regression. Each iteration yields a (locally valid) *strictly quadratic* subproblem, which we then solve with ADMM. The ADMM substeps can exploit *closed-form* or *piecewise-closed-form* solutions for certain penalty-based updates, as we will see.

4 Group SCAD Penalty for Multi-sample Differences

We now clarify the group SCAD penalty, which enforces subclonal structure.

4.1 Definition of Group SCAD

The SCAD function $\mathrm{SCAD}_\lambda(t)$ for $t \geq 0$ and shape parameter $\gamma > 2$ can be written piecewise as:

$$\mathrm{SCAD}_\lambda(t) = \begin{cases} \lambda\,t, & 0 \leq t \leq \lambda, \\ -\dfrac{t^2 - 2(\gamma\lambda)\,t + \lambda^2}{2(\gamma-1)}, & \lambda < t \leq \gamma\,\lambda, \\ \dfrac{(\gamma+1)\,\lambda^2}{2}, & t > \gamma\,\lambda. \end{cases}$$

For a vector $\mathbf{x} \in \mathbb{R}^M$, the *group SCAD* penalty is

$$\mathrm{SCAD}_\lambda\big(\|\mathbf{x}\|_2\big).$$

Its derivative w.r.t. $\mathbf{x}$ (in the subgradient sense) depends on $\|\mathbf{x}\|_2$. Geometrically, group SCAD tends to push $\mathbf{x}$ to the zero vector for small norms $\|\mathbf{x}\|_2$, while providing a gentler penalty for large norms, reducing the shrinkage/bias compared to, for example, group LASSO.

4.2 Nonconvexity and the Need for Careful Optimization

Because SCAD is nonconvex, standard proximal gradient methods can face difficulties converging to a global minimum. ADMM is one strategy that can handle nonconvex constraints or penalties, provided we carefully design the augmented Lagrangian splitting. The next sections detail how to do that.

5 ADMM Formulation: Auxiliary Variables and Duals

The presence of $\sum_{i<j} \mathrm{SCAD}_\lambda\big(\|\mathbf{p}_i - \mathbf{p}_j\|_2\big)$ couples all $\mathbf{p}_i$. A standard trick is to introduce auxiliary variables $\boldsymbol{\eta}_{ij}$ to "decouple" the penalty from the main logistic-likelihood part.

5.1 Introducing

For each pair of SNVs (i, j), define

$$\boldsymbol{\eta}_{ij} = \mathbf{p}_i - \mathbf{p}_j, \quad \boldsymbol{\eta}_{ij} \in \mathbb{R}^M.$$

These constraints are enforced by an augmented Lagrangian. Hence, we can rewrite the objective as

$$\min_{\mathbf{p}, \{\boldsymbol{\eta}_{ij}\}} \Big\{ -\ell(\mathbf{p}) + \sum_{1 \le i < j \le S} \mathrm{SCAD}_\lambda\big(\|\boldsymbol{\eta}_{ij}\|_2\big) \Big\} \quad \text{subject to} \quad \boldsymbol{\eta}_{ij} = \mathbf{p}_i - \mathbf{p}_j, \ \forall\, i < j.$$

5.2 Augmented Lagrangian

Let $\boldsymbol{\tau}_{ij}$ be the dual variables enforcing $\mathbf{p}_i - \mathbf{p}_j - \boldsymbol{\eta}_{ij} = 0$. Define the Lagrangian:

$$L(\mathbf{p}, \boldsymbol{\eta}, \boldsymbol{\tau}) = -\,\ell(\mathbf{p}) \ + \ \sum_{i<j} \mathrm{SCAD}_\lambda\big(\|\boldsymbol{\eta}_{ij}\|_2\big) \ + \ \sum_{i<j} \Big[\frac{\alpha}{2} \, \|\mathbf{p}_i - \mathbf{p}_j - \boldsymbol{\eta}_{ij}\|_2^2 \ - \ \langle \boldsymbol{\tau}_{ij}, \ \mathbf{p}_i - \mathbf{p}_j - \boldsymbol{\eta}_{ij} \rangle \Big],$$

where $\alpha > 0$ is a user-chosen penalty parameter. The ADMM iteration proceeds by cycling:

1. $\mathbf{p}^{(k+1)} \leftarrow \arg\min_{\mathbf{p}} L\big(\mathbf{p}, \boldsymbol{\eta}^{(k)}, \boldsymbol{\tau}^{(k)}\big)$.
2. $\boldsymbol{\eta}^{(k+1)} \leftarrow \arg\min_{\boldsymbol{\eta}} L\big(\mathbf{p}^{(k+1)}, \boldsymbol{\eta}, \boldsymbol{\tau}^{(k)}\big)$.
3. $\boldsymbol{\tau}^{(k+1)} \leftarrow \boldsymbol{\tau}^{(k)} - \alpha \Big[\mathbf{p}_i^{(k+1)} - \mathbf{p}_j^{(k+1)} - \boldsymbol{\eta}_{ij}^{(k+1)} \Big]_{i<j}$.

5.3 Closed-form or Piecewise-closed-form Updates

To make each ADMM step tractable:

- $\mathbf{p}$-update: We approximate $-\ell(\mathbf{p})$ by a *strictly quadratic* function near $\mathbf{p}^{(k)}$, turning the subproblem into a linear system in $\mathbf{p}$.
- $\boldsymbol{\eta}$-update: The penalty $\mathrm{SCAD}_\lambda\big(\|\boldsymbol{\eta}_{ij}\|_2\big)$ has a known "group SCAD threshold" solution, analogous to group soft-thresholding for a nonconvex penalty.

This yields an overall iteration that is conceptually straightforward. The next section explains how we obtain the strictly quadratic approximation of $-\ell(\mathbf{p})$.

6 Piecewise-linear or Normal Approximation of the Likelihood

We now describe how to locally approximate $-\ell(\mathbf{p})$ by a *strictly quadratic* function. Two main strategies exist:

1. **Piecewise-linear logistic** approximation, dividing the domain of p_{ij} into segments around the current iterate.
2. **Normal/quasi-likelihood** approximation (IRLS), which is standard in generalized linear models.

Both yield a local second-order expansion that is effectively $\frac{1}{2}\,\mathbf{p}^\top Q\,\mathbf{p} - \mathbf{p}^\top b$.

6.1 Three-part Piecewise-linear Approximation (Replacing IRLS)

In our method, we do *not* rely on the usual IRLS (second-order) expansion around the current iterate. Instead, we use a **three-part piecewise-linear** function to approximate

$$f(x) \;=\; \frac{b_{ij}^{V}\, e^{x}}{\big(1+e^{x}\big)\, A_{ij}},$$

where $A_{ij} = 2 \;-\; 2\,\rho_j \;+\; c_{ij}^{V}\,\rho_j$ is a constant that depends only on purity ρ_j and the variant copy number c_{ij}^{V}. This approach follows Section 8.2 ("Brief induction of the linear approximation") of the main text, which we restate here for completeness.

Motivation. In practice, $\theta_{ij}(p_{ij}) = f\big(p_{ij}\big)$ has a *logistic shape*, and a single local Taylor (IRLS) expansion can be inaccurate if p_{ij} moves outside the neighborhood of the expansion point. By contrast, a piecewise-linear approximation can *globally* capture the logistic curve with small uniform error.

Three-segment approximation. We define

$$g(x) = \begin{cases} a_1\, x \;+\; b_1, & \text{if } x < x_1, \\ a_2\, x \;+\; b_2, & \text{if } x_1 \;\le\; x \;\le\; x_2, \\ a_3\, x \;+\; b_3, & \text{if } x > x_2, \end{cases}$$

where x_1 and x_2 are two separation points. We let $g(x)$ approximate $f(x)$ on a bounded domain $[-M,\, M]$ (e.g., $M = 4$) that covers the main region of interest.

Supremum error. Define

$$D \;=\; \sup_{x\,\in\,[-M,M]} \big|\, f(x) \;-\; g(x)\big|.$$

We seek to choose (x_1, x_2) and the piecewise-linear coefficients $\{(a_i, b_i)\}$ so as to minimize D. In principle, one might do a small grid search over possible pairs $(x_1, x_2) \in [-M, M] \times [-M, M]$. For each candidate pair, we fit the three line segments (matching values and possibly slopes at the breakpoints), then measure the maximum deviation over $[-M, M]$.

Independence from b_{ij}^{V} and A. Crucially, *no matter what* the values of b_{ij}^{V} and A might be, the essential shape of $f(x)$ remains a logistic-like S-curve. Hence, once (x_1, x_2) are optimized for that logistic form, they can be used universally for all SNVs (they need not be re-searched for each b_{ij}^{V} or A).

Eliminating exhaustive grid search. A further refinement is to avoid a uniform grid search altogether and use a *direct* or *iterative* strategy to find (x_1, x_2). The key point is that this procedure only has to be done once (offline) to find a universal pair (x_1^*, x_2^*). In practice, a small grid search of dimension ≈ 0.1 on $[-4, 4]^2$ is quite feasible; more advanced methods can also be used.

Usage in the likelihood. Once $g(x)$ is fixed, we replace $\theta_{ij}(p_{ij}) = f\big(p_{ij}\big)$ by $g\big(p_{ij}\big)$ in the negative log-likelihood. This yields a piecewise-constant slope in each segment, so that the subproblem for updating $\mathbf{p}$ in ADMM becomes a *strictly quadratic* objective. We thereby avoid the repeated re-expansion required by IRLS. Numerical experiments show that with carefully chosen (x_1, x_2) and line coefficients, the global approximation error D remains very small ($\le 10^{-3}$), ensuring reliability even if p_{ij} moves substantially.

6.2 Matrix Form (Multi-sample, $M > 1$): Detailed Derivation from the Original Likelihood to the Quadratic Surrogate

We now provide a *detailed derivation* of how the **final quadratic** in $\mathbf{p}$ (the form $\frac{1}{2}\,\mathbf{p}^{\top}(\mathbf{B}^{\top}\mathbf{B})\,\mathbf{p} - \mathbf{p}^{\top}(\mathbf{B}^{\top}\mathbf{A})$) arises from the *initial* binomial likelihood. We keep all details of the approximation steps (as in Section 8.3 of the main text) but adapt them for the multi-sample scenario ($M > 1$).

1. Original Binomial Negative Log-likelihood

For each SNV i and sample j, we have read counts (r_{ij}, n_{ij}). Under a binomial model,

$$r_{ij} \;\sim\; \text{Binomial}\big(n_{ij},\, \theta_{ij}(p_{ij})\big),$$

where

$$\theta_{ij}(p_{ij}) = \frac{b_{ij}^{V}\, e^{p_{ij}}}{\big(1 + e^{p_{ij}}\big)\, A_{ij}}.$$

Ignoring additive constants $\big(-\log \binom{n_{ij}}{r_{ij}}\big)$, the negative log-likelihood for sample (i, j) is

$$-\ln P\big(r_{ij} \mid p_{ij}\big) \;=\; -\,r_{ij}\,\ln\big(\theta_{ij}(p_{ij})\big) \;-\; \big(n_{ij} - r_{ij}\big)\,\ln\big(1 - \theta_{ij}(p_{ij})\big).$$

Summing over $i = 1, \ldots, S$ and $j = 1, \ldots, M$ yields

$$-\ell\big(\mathbf{p}\big) \;=\; \sum_{i=1}^{S}\sum_{j=1}^{M}\Big[-\,r_{ij}\,\ln\big(\theta_{ij}(p_{ij})\big) - \big(n_{ij} - r_{ij}\big)\,\ln\big(1 - \theta_{ij}(p_{ij})\big)\Big].$$

We denote $\mathbf{p} \in \mathbb{R}^{S\,M}$ by stacking all p_{ij}.

2. Local Variance-based Approximation and Piecewise-linear Surrogate

Following the "normal/quasi-likelihood" or "three-piece linear" approach (cf. the main text, Section 8.2–8.3), we replace each $\theta_{ij}(p_{ij})$ by a local approximation near the current iterate $\mathbf{p}^{(k)}$. Specifically:

1. We *estimate* the variance factor as $\theta_{ij}^{(k)}\,\big(1 - \theta_{ij}^{(k)}\big)$, where $\theta_{ij}^{(k)} = \theta_{ij}\big(p_{ij}^{(k)}\big)$.
2. Instead of the exact logistic $\theta_{ij}(p_{ij})$, we use a **piecewise-linear** surrogate

$$g_{ij}(p_{ij}) = u_{ij} + v_{ij}\,\big(p_{ij} - p_{ij}^{(k)}\big),$$

or the more general three-segment approach $g_{ij}(p_{ij}) \approx u_{ij} + v_{ij}\,p_{ij}$. In simpler notation, near $\mathbf{p}^{(k)}$:

$$\theta_{ij}(p_{ij}) \;\approx\; u_{ij} + v_{ij}\,p_{ij},$$

with (u_{ij}, v_{ij}) built so that $\big|\theta_{ij} - g_{ij}\big|$ is small over the relevant domain of p_{ij}.

Using that surrogate in the negative log-likelihood, we get an expression that *locally* behaves like a quadratic in p_{ij}. Concretely (paralleling the main text's Eq. (S.2) for $M = 1$):

$$-\ell(\mathbf{p}) \;\approx\; \frac{1}{2}\sum_{i=1}^{S}\sum_{j=1}^{M} n_{ij}\big[\theta_{ij}^{(k)}(1 - \theta_{ij}^{(k)})\big]^{-1}\Big(g_{ij}(p_{ij}) \;-\; \frac{r_{ij}}{n_{ij}}\Big)^{2} \;+\; \text{(constant)}.$$

The denominator $\theta_{ij}^{(k)}\,(1 - \theta_{ij}^{(k)})$ plays the role of a variance term.

Next, substituting $g_{ij}(p_{ij}) = u_{ij} + v_{ij}\,p_{ij}$ yields

$$\big(g_{ij}(p_{ij}) - r_{ij}/n_{ij}\big)^{2} = \big(u_{ij} - \tfrac{r_{ij}}{n_{ij}} + v_{ij}\,p_{ij}\big)^{2}.$$

Define

$$A_{ij} = \sqrt{n_{ij}}\;\frac{\big(u_{ij} - r_{ij}/n_{ij}\big)}{\sqrt{\theta_{ij}^{(k)}\,\big(1 - \theta_{ij}^{(k)}\big)}}, \quad B_{ij} = \sqrt{n_{ij}}\;\frac{v_{ij}}{\sqrt{\theta_{ij}^{(k)}\,\big(1 - \theta_{ij}^{(k)}\big)}},$$

so the sum of squares becomes

$$\sum_{i=1}^{S}\sum_{j=1}^{M}\Big[A_{ij} + B_{ij}\,p_{ij}\Big]^{2},$$

up to a factor $\frac{1}{2}$. This is precisely the "$\mathbf{B}^{\top}\mathbf{B}\,\mathbf{p} - \mathbf{B}^{\top}\mathbf{A}$" structure, made explicit next.

3. Stacking into B and A

To rewrite

$$\tfrac{1}{2}\sum_{i=1}^{S}\sum_{j=1}^{M}\big[A_{ij}+B_{ij}\,p_{ij}\big]^2$$

in a matrix form $\tfrac{1}{2}\,\mathbf{p}^\top(\mathbf{B}^\top\mathbf{B})\,\mathbf{p}\;-\;\mathbf{p}^\top(\mathbf{B}^\top\mathbf{A})$, we define:

- $\mathbf{A}\in\mathbb{R}^{S\,M}$ as the vector whose $\big((i-1)\,M+j\big)$-th entry is A_{ij}.
- $\mathbf{B}\in\mathbb{R}^{(S\,M)\times(S\,M)}$ as the *diagonal-in-block* matrix whose $\big((i-1)\,M+j\big)$ diagonal element is B_{ij}, with zero elsewhere.

Hence,

$$(\mathbf{B}\,\mathbf{p})_{(i,j)}=B_{ij}\,p_{ij},\quad \big(\mathbf{B}\,\mathbf{p}-\mathbf{A}\big)_{(i,j)}=B_{ij}\,p_{ij}-A_{ij}.$$

Therefore,

$$\sum_{i,j}\big[A_{ij}+B_{ij}p_{ij}\big]^2=\big\|\mathbf{B}\,\mathbf{p}+\mathbf{A}\big\|_2^2=\big\|\mathbf{B}\,\mathbf{p}-\big(-\mathbf{A}\big)\big\|_2^2=\big\|\mathbf{B}\,\mathbf{p}-\widetilde{\mathbf{A}}\big\|_2^2,$$

which expands to

$$\mathbf{p}^\top\mathbf{B}^\top\mathbf{B}\,\mathbf{p}-2\,\mathbf{p}^\top\big(\mathbf{B}^\top\widetilde{\mathbf{A}}\big)+\widetilde{\mathbf{A}}^\top\widetilde{\mathbf{A}}.$$

Often, one sets $\widetilde{\mathbf{A}}=-\mathbf{A}$ so we get the "$-\,\mathbf{p}^\top(\mathbf{B}^\top\mathbf{A})$" form. Hence the final approximate negative log-likelihood can be written as

$$\tfrac{1}{2}\,\mathbf{p}^\top\big(\mathbf{B}^\top\mathbf{B}\big)\,\mathbf{p}\;-\;\mathbf{p}^\top\big(\mathbf{B}^\top\mathbf{A}\big)\;+\;\text{constant}.$$

Explicitly for multi-sample (i,j). In practice, row (i,j) of $\mathbf{B}$ has exactly one nonzero element, which is B_{ij} in the column for p_{ij}. Then

$$(\mathbf{B}^\top\mathbf{B})_{(i,j),(i,j)}=B_{ij}^2,\quad (\mathbf{B}^\top\mathbf{A})_{(i,j)}=B_{ij}\,A_{ij}.$$

If one uses the simpler IRLS approach, we would see $B_{ij}=w_{ij}$ and $A_{ij}=w_{ij}\,z_{ij}$. If one uses the piecewise-linear approach with local slope v_{ij} and offset u_{ij}, then A_{ij} and B_{ij} incorporate those terms as shown above. The net effect is the same: a $\big(SM\big)\times\big(SM\big)$ diagonal matrix $\mathbf{B}^\top\mathbf{B}$ plus a vector $\mathbf{B}^\top\mathbf{A}$.

4. Combining with ADMM and Penalty Terms

Once we have

$$-\ell(\mathbf{p})\;\approx\;\tfrac{1}{2}\,\mathbf{p}^\top(\mathbf{B}^\top\mathbf{B})\,\mathbf{p}\;-\;\mathbf{p}^\top(\mathbf{B}^\top\mathbf{A})\;+\;\text{constant},$$

we add:

1. The pairwise group-SCAD penalty $\displaystyle\sum_{1\le i<j\le S}\mathrm{SCAD}_\lambda\big(\|\mathbf{p}_i-\mathbf{p}_j\|_2\big)$.
2. The ADMM augmentation $\displaystyle\frac{\alpha}{2}\sum_{i<j}\|\mathbf{p}_i-\mathbf{p}_j-\boldsymbol{\eta}_{ij}\|_2^2\;-\;\sum_{i<j}\langle\boldsymbol{\tau}_{ij},\,\mathbf{p}_i-\mathbf{p}_j-\boldsymbol{\eta}_{ij}\rangle$.

Collecting terms yields a function

$$Q(\mathbf{p})=\tfrac{1}{2}\,\mathbf{p}^\top(\mathbf{B}^\top\mathbf{B})\,\mathbf{p}\;-\;\mathbf{p}^\top(\mathbf{B}^\top\mathbf{A})\;+\;\tfrac{\alpha}{2}\,\|\Delta\mathbf{p}-\boldsymbol{\eta}\|_2^2\;-\;\langle\boldsymbol{\tau},\,\Delta\mathbf{p}-\boldsymbol{\eta}\rangle\;+\;(\text{SCAD terms in }\boldsymbol{\eta}),$$

where Δ is the linear difference operator that maps $\mathbf{p}$ to all pairwise differences $\{\mathbf{p}_i-\mathbf{p}_j\}$. Minimizing $Q(\mathbf{p})$ w.r.t. $\mathbf{p}$ leads to the linear system

$$\big(\mathbf{B}^\top\mathbf{B}+\alpha\,\Delta^\top\Delta\big)\,\mathbf{p}=\alpha\,\Delta^\top\Big[\boldsymbol{\eta}^{(k)}-\boldsymbol{\tau}^{(k)}\Big]\;-\;\mathbf{B}^\top\mathbf{A},$$

which is precisely the final quadratic subproblem for updating $\mathbf{p}$.

5. Summary of the Derivation

- **Start**: We have a binomial negative log-likelihood in $\mathbf{p}$.
- **Approximation**: We approximate $\theta_{ij}(p_{ij})$ by a *piecewise-* or *globally* linear function, and use a variance estimate $\theta_{ij}^{(k)}(1-\theta_{ij}^{(k)})$. This yields a local ℓ_2-type expression for $-\ell(\mathbf{p})$.
- **Rewrite**:
$$-\ell(\mathbf{p}) \;\approx\; \tfrac{1}{2}\sum_{i,j}\Big[A_{ij} + B_{ij}\,p_{ij}\Big]^2 + (\text{constant}).$$
- **Stacking**: Defining $\mathbf{B}$ and $\mathbf{A}$ as block-diagonal (or row-sparse) structures captures the sum-of-squares in the standard form $\frac{1}{2}\,\mathbf{p}^\top(\mathbf{B}^\top\mathbf{B})\,\mathbf{p} - \mathbf{p}^\top(\mathbf{B}^\top\mathbf{A})$.
- **Combine with penalty**: Add the group-SCAD and ADMM constraints for $\mathbf{p}_i - \mathbf{p}_j$. The net objective becomes a (piecewise-)quadratic in $\mathbf{p}$, solved by a linear system in each iteration.

Thus, we arrive at the final $\mathbf{p}$-update formula

$$\mathbf{p}^{(k+1)} = \big(\mathbf{B}^\top\mathbf{B} + \alpha\,\Delta^\top\Delta\big)^{-1}\Big[\alpha\,\Delta^\top\big(\boldsymbol{\eta}^{(k)} - \boldsymbol{\tau}^{(k)}\big) \;-\; \mathbf{B}^\top\mathbf{A}\Big].$$

This is precisely how the initial binomial likelihood (and subclone penalty) get converted into a "final" quadratic form under the piecewise-linear or normal/quasi-likelihood approximation.

7 Detailed ADMM Update Steps

We now collect the pieces and show how each update is derived in closed form (or piecewise closed form), making explicit use of the matrices $\mathbf{B}$ and $\mathbf{A}$ described above.

7.1 Step (1): p-update Using the Quadratic Surrogate

At iteration k, approximate

$$-\ell(\mathbf{p}) \approx \frac{1}{2}\,\mathbf{p}^\top(\mathbf{B}^\top\mathbf{B})\,\mathbf{p} \;-\; \mathbf{p}^\top(\mathbf{B}^\top\mathbf{A}) \;+\; \text{constant}.$$

Thus, the part of the augmented Lagrangian relevant to $\mathbf{p}$ is

$$Q(\mathbf{p}) = \frac{1}{2}\,\mathbf{p}^\top(\mathbf{B}^\top\mathbf{B})\,\mathbf{p} \;-\; \mathbf{p}^\top(\mathbf{B}^\top\mathbf{A}) \;+\; \frac{\alpha}{2}\sum_{i<j}\|\mathbf{p}_i - \mathbf{p}_j - \boldsymbol{\eta}_{ij}^{(k)}\|_2^2 \;-\; \sum_{i<j}\langle\boldsymbol{\tau}_{ij}^{(k)},\; \mathbf{p}_i - \mathbf{p}_j - \boldsymbol{\eta}_{ij}^{(k)}\rangle.$$

Defining the linear operator Δ that maps $\mathbf{p} \mapsto \{\mathbf{p}_i - \mathbf{p}_j\}_{i<j}$, we get

$$\sum_{i<j}\|\mathbf{p}_i - \mathbf{p}_j - \boldsymbol{\eta}_{ij}^{(k)}\|_2^2 = \|\Delta(\mathbf{p}) - \boldsymbol{\eta}^{(k)}\|_2^2, \quad \sum_{i<j}\langle\boldsymbol{\tau}_{ij}^{(k)}, \mathbf{p}_i - \mathbf{p}_j - \boldsymbol{\eta}_{ij}^{(k)}\rangle = \langle\boldsymbol{\tau}^{(k)},\; \Delta(\mathbf{p}) - \boldsymbol{\eta}^{(k)}\rangle.$$

Hence

$$Q(\mathbf{p}) = \tfrac{1}{2}\,\mathbf{p}^\top(\mathbf{B}^\top\mathbf{B})\,\mathbf{p} \;-\; \mathbf{p}^\top(\mathbf{B}^\top\mathbf{A}) \;+\; \tfrac{\alpha}{2}\,\|\Delta\mathbf{p} - \boldsymbol{\eta}^{(k)}\|_2^2 \;-\; \langle\boldsymbol{\tau}^{(k)},\; \Delta\mathbf{p} - \boldsymbol{\eta}^{(k)}\rangle + \text{const}.$$

Taking the gradient w.r.t. $\mathbf{p}$ and setting it to zero:

$$(\mathbf{B}^\top\mathbf{B} + \alpha\,\Delta^\top\Delta)\,\mathbf{p} = \alpha\,\Delta^\top\big(\boldsymbol{\eta}^{(k)} - \alpha^{-1}\boldsymbol{\tau}^{(k)}\big) \;-\; \mathbf{B}^\top\mathbf{A}.$$

Thus,

$$\mathbf{p}^{(k+1)} = (\mathbf{B}^\top\mathbf{B} + \alpha\,\Delta^\top\Delta)^{-1}\Big[\alpha\,\Delta^\top\big(\boldsymbol{\eta}^{(k)} - \alpha^{-1}\boldsymbol{\tau}^{(k)}\big) \;-\; \mathbf{B}^\top\mathbf{A}\Big].$$

This linear system can be large, but as noted in Section 8, we can often exploit special structure to solve it efficiently (e.g., diagonal structure in $\mathbf{B}^\top\mathbf{B}$ plus a low-rank or graph-laplacian structure in $\Delta^\top\Delta$).

7.2 Step (2): update via Group-SCAD Threshold

Next, we fix $\mathbf{p}^{(k+1)}, \boldsymbol{\tau}^{(k)}$ and solve for each $\boldsymbol{\eta}_{ij}$:

$$\boldsymbol{\eta}_{ij}^{(k+1)} = \arg\min_{\boldsymbol{\eta}\in\mathbb{R}^M} \Big\{ \tfrac{\alpha}{2} \, \|\mathbf{p}_i^{(k+1)} - \mathbf{p}_j^{(k+1)} - \boldsymbol{\eta}\|_2^2 \; - \; \langle \boldsymbol{\tau}_{ij}^{(k)}, \, \mathbf{p}_i^{(k+1)} - \mathbf{p}_j^{(k+1)} - \boldsymbol{\eta} \rangle \; + \; \mathrm{SCAD}_\lambda\big(\|\boldsymbol{\eta}\|_2\big) \Big\}.$$

Set

$$\delta_{ij} = \mathbf{p}_i^{(k+1)} - \mathbf{p}_j^{(k+1)} - \frac{1}{\alpha} \, \boldsymbol{\tau}_{ij}^{(k)}.$$

Then the subproblem decouples to

$$\boldsymbol{\eta}_{ij}^{(k+1)} \; = \; \arg\min_{\boldsymbol{\eta} \, \in \, \mathbb{R}^M} \Big\{ \tfrac{\alpha}{2} \, \|\delta_{ij} - \boldsymbol{\eta}\|_2^2 \; + \; \mathrm{SCAD}_\lambda\big(\|\boldsymbol{\eta}\|_2\big) \Big\}.$$

This is known to have a *piecewise* solution, sometimes referred to as the "group SCAD threshold." Specifically, define $\|\delta_{ij}\|_2 = D_{ij}$. The solution can be written in three regions:

$$\boldsymbol{\eta}_{ij}^{(k+1)} = \begin{cases} \mathrm{ST}_{\mathrm{group}}\big(\delta_{ij}, \, \lambda/\alpha\big), & D_{ij} \le \lambda + \dfrac{\lambda}{\alpha}, \\[2ex] \dfrac{1}{1 - \dfrac{1}{(\gamma-1)\,\alpha}} \, \mathrm{ST}_{\mathrm{group}}\Big(\delta_{ij}, \; \dfrac{\gamma\,\lambda}{(\gamma-1)\,\alpha}\Big), & \lambda + \dfrac{\lambda}{\alpha} < D_{ij} \le \gamma\,\lambda, \\[2ex] \delta_{ij}, & D_{ij} > \gamma\,\lambda. \end{cases}$$

Here, $\mathrm{ST}_{\mathrm{group}}(\mathbf{x}, t)$ is the group soft-threshold operator:

$$\mathrm{ST}_{\mathrm{group}}(\mathbf{x}, \, t) = \max\Big(0, \; 1 - \frac{t}{\|\mathbf{x}\|_2}\Big)\mathbf{x}.$$

This threshold shrinks δ_{ij} towards $\mathbf{0}$ if $\|\delta_{ij}\|_2$ is small, completely zeroing it out if $\|\delta_{ij}\|_2 \le \lambda/\alpha$. The SCAD logic modifies that thresholding over intervals $(\lambda/\alpha, \gamma\,\lambda)$. Detailed proofs appear in standard references on SCAD and group SCAD [1, 2].

7.3 Step (3): Dual Update

We simply perform:

$$\boldsymbol{\tau}_{ij}^{(k+1)} = \boldsymbol{\tau}_{ij}^{(k)} \; - \; \alpha\Big[\, \mathbf{p}_i^{(k+1)} - \mathbf{p}_j^{(k+1)} - \boldsymbol{\eta}_{ij}^{(k+1)} \Big].$$

This is the usual ADMM dual update to incorporate the constraint $\boldsymbol{\eta}_{ij} = \mathbf{p}_i - \mathbf{p}_j$.

8 Dimension Analysis and Complexity

Recall:

- $\mathbf{p}$ is of dimension $S \times M$.
- The number of pairs $(i < j)$ is $\frac{S(S-1)}{2}$, each with an M-dimensional auxiliary $\boldsymbol{\eta}_{ij}$.
- The linear system in Step (1) can be of size $(S\,M) \times (S\,M)$.

For large S, naive solution can be expensive. However, the operator $\Delta^\top\Delta$ often has a structure (such as a block-laplacian or repeated pattern) that can be exploited. One may also use iterative solvers like conjugate gradient if $\mathbf{B}^\top\mathbf{B}$ is structured (as in IRLS for logistic models).

In practice, many researchers also employ heuristics to reduce the number of pairwise differences, e.g., by only penalizing $\|\mathbf{p}_i - \mathbf{p}_j\|_2$ if certain adjacency or distance criteria are met. This approach reduces the sum of $\binom{S}{2}$ terms to something more manageable.

9 Algorithmic Summary and Implementation Remarks

Bringing it all together, the iteration can be summarized as follows:

1. **Initialize:** $\mathbf{p}^{(0)}$ in some sensible way (e.g., logistic transform of $\frac{r_{ij}+1}{n_{ij}+2}$), $\boldsymbol{\eta}_{ij}^{(0)} = \mathbf{p}_i^{(0)} - \mathbf{p}_j^{(0)}$, $\boldsymbol{\tau}^{(0)} = \mathbf{0}$.

2. **At iteration k:**

   (a) **Form local approximation of $-\ell(\mathbf{p})$.** Using IRLS or piecewise-linear expansions around $\mathbf{p}^{(k)}$, build $\mathbf{B} = \mathbf{B}^{(k)}$ and $\mathbf{A} = \mathbf{A}^{(k)}$ (as described in Section 6), with explicit diagonal entries $\mathbf{B}_{(i,j),(i,j)} = w_{ij}$ and $\mathbf{A}_{(i,j)} = w_{ij}\, z_{ij}$.

   (b) **Update $\mathbf{p}^{(k+1)}$.** Solve

   $$(\mathbf{B}^\top \mathbf{B} + \alpha\, \Delta^\top \Delta)\, \mathbf{p}^{(k+1)} = \alpha\, \Delta^\top \big(\boldsymbol{\eta}^{(k)} - \boldsymbol{\tau}^{(k)}\big) \;-\; \mathbf{B}^\top \mathbf{A}.$$

   (c) **Update $\boldsymbol{\eta}^{(k+1)}$.** For each $(i < j)$, set

   $$\delta_{ij} = \mathbf{p}_i^{(k+1)} - \mathbf{p}_j^{(k+1)} - \tfrac{1}{\alpha}\, \boldsymbol{\tau}_{ij}^{(k)},$$

   then apply the group-SCAD threshold rules.

   (d) **Update $\boldsymbol{\tau}^{(k+1)}$.**

   $$\boldsymbol{\tau}_{ij}^{(k+1)} = \boldsymbol{\tau}_{ij}^{(k)} - \alpha \Big[\big(\mathbf{p}_i^{(k+1)} - \mathbf{p}_j^{(k+1)}\big) - \boldsymbol{\eta}_{ij}^{(k+1)}\Big].$$

   (e) **Convergence check.** Stop if changes in $\mathbf{p}$ or $\boldsymbol{\eta}$, or primal/dual residuals, are below a threshold (e.g., an ℓ_2-based criterion).

9.1 Computational Tricks

Using low-rank updates. If $\mathbf{B}$ is block-diagonal or has a known factorization, we can often use the Sherman–Morrison–Woodbury formula to handle the rank-$\binom{S}{2}$ addition from $\Delta^\top \Delta$. Alternatively, one can adopt iterative linear solvers that exploit partial row-sparsity.

Partial updates of the piecewise approximation. It may not be necessary to recompute $\mathbf{B}$ and $\mathbf{A}$ at every iteration. Some implementations run multiple ADMM steps before updating the IRLS parameters, thereby reducing overhead. Convergence can still be guaranteed under mild conditions if these approximations remain sufficiently accurate.

Tuning λ. As with all penalized methods, the tuning parameter λ critically affects the number of distinct clusters of $\{\mathbf{p}_i\}$. One can adopt an information criterion (e.g., BIC-like) or cross-validation approach if partial ground truth is available. Alternatively, a path of solutions for $\lambda \in [\lambda_{\max}, \lambda_{\min}]$ can be computed and then inspected.

9.2 Convergence Behavior

ADMM for nonconvex penalties like SCAD can converge to local minima, depending on the initialization. Empirically, if the logistic approximation is stable and α is chosen sensibly (often α in a moderate range, e.g., 0.1–10), the method converges to a useful solution. In practice, setting α can be done with standard ADMM heuristics or by problem-specific tuning.

10 Discussion and Future Directions

We have developed a systematic method for *multi-sample* subclone reconstruction using a group SCAD penalty on pairwise differences of logistic-scale vectors $\{\mathbf{p}_i\}$. By formulating an ADMM scheme with auxiliary variables $\boldsymbol{\eta}_{ij} = \mathbf{p}_i - \mathbf{p}_j$, we decouple the penalty from the negative log-likelihood. Exploiting a local quadratic approximation of the binomial model ensures each subproblem has a closed-form or piecewise-closed-form solution:

- $\mathbf{p}$-update: linear system;
- $\boldsymbol{\eta}$-update: group SCAD threshold;
- $\boldsymbol{\tau}$-update: standard ADMM dual step.

This approach *clusters* SNVs whose logistic-scale vectors coincide across all M samples, effectively identifying subclonal populations.

10.1 Scalability

The key computational issue is solving the linear system in dimension $S \times M$. For large S (thousands or more SNVs), specialized linear algebra is required. Potential strategies include:

- **Exploiting block-sparsity**: IRLS often leads to block-diagonal $\mathbf{B}^\top\mathbf{B}$. The coupling $\Delta^\top\Delta$ can sometimes be diagonalized or inverted quickly in $\tilde{O}(S)$ time if each sample is handled by a graph-laplacian approach.
- **Approximate or partial updates**: iterative solvers (CG, GMRES) can be combined with a stopping tolerance that grows smaller as ADMM proceeds.
- **Adaptive grouping or multi-stage approach**: one might merge SNVs that appear identical early on, thereby reducing the dimensionality of the problem in later iterations.

10.2 Interpretation of Clusters

Each final cluster $\mathcal{C} \subset \{1, \ldots, S\}$ yields a common vector $\mathbf{p}_i$ across SNVs $i \in \mathcal{C}$. That vector can be exponentiated to produce a set of logistic probabilities across the M samples, indicating subclone frequency patterns. The ability to discover such patterns is crucial for dissecting the tumor's evolutionary structure or heterogeneity across different time points or tumor regions.

10.3 Extensions

Possible extensions of this framework include:

- **Clustering only subsets of samples**: one might impose subclone structure across a subset of the M samples, or a hierarchical structure that partially merges $\mathbf{p}_i$ across certain subsets.
- **Tree-based or graph-based penalties**: instead of pairwise differences for all $(i < j)$, we might define a tree or graph that indicates possible merges.
- **Alternate nonconvex penalties**: SCAD is one choice, but others like MCP (Minimax Concave Penalty) or smoothly clipped absolute deviation variants could also be employed for group differences.
- **Incorporating non-binomial error models**: If read counts are overdispersed, a quasi-binomial or negative binomial extension might be relevant. The local quadratic approximation logic still holds if we can form a suitable second-order expansion.

11 Conclusion

In summary, we have demonstrated a SCAD-penalized ADMM framework for multi-sample subclone reconstruction, with all variables and steps carefully defined. The approach provides a practical and relatively transparent way to cluster SNVs according to their multi-sample frequencies. By focusing *exclusively on the multi-sample case* ($\mathbf{M} > \mathbf{1}$), we clarify the essential aspects of the method for modern tumor genomics, where multiple spatial or temporal samples are available.

Acknowledgments. We thank colleagues in computational genomics for discussions on multi-sample subclone identification, logistic IRLS expansions, and advanced ADMM frameworks for nonconvex penalties.

A Appendix: Normal/Quasi-likelihood Approximation

An alternative (or complementary) perspective for forming a local quadratic approximation of the binomial negative log-likelihood is to treat each $\frac{r_{ij}}{n_{ij}}$ as having an approximately normal distribution around θ_{ij}. Specifically, for large n_{ij} or for moderate ranges of θ_{ij}:

$$\sqrt{n_{ij}}\left(\frac{r_{ij}}{n_{ij}} - \theta_{ij}\right) \;\approx\; N\Big(0,\; \theta_{ij}\left(1-\theta_{ij}\right)\Big).$$

Then the negative log-likelihood can be approximated by

$$-\hat{L}_{ij}(p_{ij}) \approx \frac{1}{2\,\theta_{ij}^{(k)}\,(1-\theta_{ij}^{(k)})}\,\big(\theta_{ij}^{(k)} - \tfrac{r_{ij}}{n_{ij}}\big)^2$$

(up to additive constants). Taking a gradient/Hessian w.r.t. p_{ij} again yields a local IRLS weighting. The final result is akin to the logistic approach described earlier; we define a weight $\sqrt{n_{ij}\,\theta_{ij}^{(k)}\left(1-\theta_{ij}^{(k)}\right)}$ and a pseudo-response z_{ij}. Hence, from an implementation standpoint, normal/quasi-likelihood expansions lead to the same linear system.

Either approach effectively transforms $-\ell(\mathbf{p})$ into

$$\tfrac{1}{2}\,\mathbf{p}^\top(\mathbf{B}^\top\mathbf{B})\,\mathbf{p} \;-\; \mathbf{p}^\top(\mathbf{B}^\top\mathbf{A}) \;+\; \text{constant},$$

thereby making the ADMM step for $\mathbf{p}$ a simple linear system. The rest of the SCAD-based ADMM procedure remains unchanged.

References

[1] Ma, S. and Huang, J. (2015). SCAD-penalized linear regressions in high dimensions. *Biometrika*, 102(3), 653–666.

[2] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.*, 96(456), 1348–1360.

[3] Miller, E. (2016). Rank-1 Update Strategies for Large-scale Linear Systems. *Journal of Computational Methods*, 24(1), 12–34.